NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

THESIS

AN ASSESSMENT OF THE IMPACT OF FUSED MONOCHROME AND FUSED COLOR NIGHT VISION DISPLAYS ON REACTION TIME AND ACCURACY IN TARGET DETECTION

by

Matthew Thomas Sampson

September, 1996

Thesis Advisor: William K. Krebs
Co-Advisor: Robert Read

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 1996.	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE
AN ASSESSMENT OF THE IMPACT OF FUSED MONOCHROME AND FUSED COLOR NIGHT VISION DISPLAYS ON REACTION TIME AND ACCURACY IN TARGET DETECTION

5. FUNDING NUMBERS

6. AUTHOR(S) Matthew Thomas Sampson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

Night Vision Devices (NVDs) employed by the military fall into two categories: Image Intensifiers (I^2) also known as Night Vision Goggles (NVGs) and Infrared (IR). Each sensor provides unique visual information not available to the unaided human visual system. However, these devices have limitations and they have been listed as a causal factor in many crashes of military aircraft at night. Researchers hypothesize that digitally fusing the output from these sensors into one image and then artificially coloring the image will improve an NVD user's visual performance. The purpose of this thesis was to determine if fusion and coloring of static, natural scene NVG and IR imagery will improve reaction time and accuracy in target detection.

Pairs of static images from three different scenes were obtained simultaneously from NVG and IR sensors. The six original images were fused pixel by pixel and then colored using a computer algorithm. A natural target was moved to two other coherent positions in the scene or completely removed, resulting in twenty-four images for each of the three natural scenes. Six subjects viewed the images randomly on a high-resolution monitor, rapidly indicating on a keypad if the target was present (1) or absent (2). Reaction time and accuracy were recorded. An ANOVA on the output and a subsequent review of the images revealed that fusion significantly impacted local (target) contrast and that, coupled with scene content, decreased performance on the task. Fusion and coloring results were not superior here, which differed from results on other types of tasks, however, more research is needed to completely assess this technology.

14. SUBJECT TERMS *Sensor fusion, fused monochrome displays, fused color displays, Human Factors, Visual Search, Target Detecion, NVG, IR.	15. NUMBER OF PAGES 129
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18 298-102

Approved for public release; distribution is unlimited.

AN ASSESSMENT OF THE IMPACT OF FUSED MONOCHROME AND FUSED COLOR NIGHT VISION DISPLAYS ON REACTION TIME AND ACCURACY IN TARGET DETECTION

Matthew T. Sampson
Captain, United States Marine Corps
B.S., United States Naval Academy, 1987

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1996

ABSTRACT

Night Vision Devices (NVDs) employed by the military fall into two categories: Image Intensifiers (I^2) also known as Night Vision Goggles (NVGs) and Infrared (IR). Each sensor provides unique visual information not available to the unaided human visual system. However, these devices have limitations and they have been listed as a causal factor in many crashes of military aircraft at night. Researchers hypothesize that digitally fusing the output from these sensors into one image and then artificially coloring the image will improve an NVD user's visual performance. The purpose of this thesis was to determine if fusion and coloring of static, natural scene NVG and IR imagery will improve reaction time and accuracy in target detection.

Pairs of static images from three different scenes were obtained simultaneously from NVG and IR sensors. The six original images were fused pixel by pixel and then colored using a computer algorithm. A natural target was moved to two other coherent positions in the scene or completely removed, resulting in twenty-four images for each of the three natural scenes. Six subjects viewed the images randomly on a high-resolution monitor, rapidly indicating on a keypad if the target was present (1) or absent (2). Reaction time and accuracy were recorded. An ANOVA on the output and a subsequent review of the images revealed that fusion significantly impacted local (target) contrast and that, coupled with scene content, decreased performance on the task. Fusion and coloring results were not superior here, which differed from results on other types of tasks, however, more research is needed to completely assess this technology.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

This thesis would not have been possible without the support of my wife, Victoria, my sons, Matthew and John and our child to be. To them I am indebted for their understanding and love in this trying time.

Special thanks to my primary advisor, Prof. "Kip" Krebs, for his patience in going above and beyond the Operations Research curriculum to teach me the intricacies of vision research needed for this thesis.

Special thanks to Dr. Dean Scribner of the Naval Research Lab, Optical Sciences Division, for his ideas, images and help in understanding the color fusion process.

Special thanks to the U.S. Army Night Vision Laboratory, Night Vision Sensor Electro-Optical Directorate for their images.

Special thanks to the Sensory Sciences Division of NAMRL, Pensacola for the hardware used in the experiment for this thesis.

I. INTRODUCTION

> "Darkness is a double edged sword, and like the terrain, it favors the one who best uses it and it hinders the one who does not." A field marshal of the former Soviet Union

A. BACKGROUND

The element of surprise has historically been one of the greatest advantages a military leader can gain over an enemy. Leaders of military ground-forces have sought the favorable edge of darkness to surprise their enemies by advancing, repositioning or removing troops in a battle area. After the dawn of military aviation and starting with World War II, U.S. military doctrine included night delivery of weapons and troops as methods to surprise the enemy. Other military leaders, like those of the Viet Cong and the North Vietnamese Army were masters at conducting night operations for insurgencies and frontal attacks on the isolated fire bases and base camps of U.S. forces in South Vietnam.

U.S. military leaders in Vietnam first tried to deny the use of darkness to the enemy with searchlights, a move that did more to pinpoint the exact location of U.S. forces. The next attempt at denial was with near infrared searchlights coupled with near-infrared viewers, the viewers being so simple and accessible that the enemy soon had them to pinpoint the location of U.S. sources. (MAWTS-1, 1995) Another technology was needed, one which could passively (without emitting energy) provide ground forces with a picture of the enemy operating near them.

Despite the existence of passive airborne sensors of the longer wavelength, far

infrared (IR) spectrum, this technology had not evolved sufficiently to provide a man-portable IR sensor to U.S. forces in Viet Nam. (Lloyd, 1975) The ultimate break-through came in the form of a passive image intensifier (I^2) tube, a more complex system than the ones already tried, but one which was man-portable and which provided the user an image from intensified ambient and reflected light.

Since the advent of I^2 devices, the U.S. military has adapted them for use by all forces. Also, IR technology has improved dramatically since 1965 such that there are currently numerous forward looking infrared (FLIR) systems in the military inventory. The first employment of a FLIR for navigation (NAVFLIR) was on the army's AH-64 Apache helicopter in the late 1970's. Today, I^2 and FLIR devices are collectively referred to by the military as night vision devices (NVDs). Some common employments of NVDs today are night vision goggles (NVGs) by infantry, aviation and naval forces, night vision ('starlight') rifle scopes by infantry units, forward looking infrared (FLIR) by aviation units and thermal (IR) targeting sights by armor and aviation units. For the purpose of this thesis, only aviation variants of these NVD will be referenced.

Because a human's perception of their surroundings at night is normally devoid of NVD imagery, NVDs have been (somewhat naively) championed as tools that virtually "turn night into day." This couldn't be farther from the truth. Despite the vast improvements in NVD technology and training, there have been users whose lack of understanding of the highly dynamic night environment and its impact on their particular NVD's performance has caused them to exceed the capabilities of these devices. Their actions have often resulted

in dire consequences. For example, NVG's have been indirectly related to several 'class A' mishaps[1]. From 1973-1993 naval aviation (USN/USMC) has incurred 13 rotary-wing and 5 fixed-wing class A mishaps while employing NVG's, resulting in 15 rotary-wing aircraft, 6 fixed-wing aircraft, and 39 lives lost. Because IR systems are primarily relied upon to assist in navigation and targeting for aircraft operating at higher altitudes (greater than 500 feet), few mishaps have FLIR listed as a causal factor.

Despite any drawbacks aircrewmen may encounter with NVDs, these systems are considered essential tools for conducting successful night operations. Reliance on NVDs for night operations is evident by both tactical fixed and rotary-wing squadron training and readiness focus shifting almost entirely toward 'aided' (NVD) operations, leaving only a few familiarization flights for 'unaided' flight. Steady improvements in NVD technology have motivated aviation forces to seek innovative ways to increase the scope of their use, which in turn has enabled capabilities validated in training to 'bubble-up' and drive night operations doctrine. By employing NVDs in ways its former enemies never dreamed of, U.S. military doctrine has evolved from strictly defensive operations at night toward a true '24-hour' battlefield. As Iraqi forces recently learned, U.S. forces are capable of 'shooting and moving' anywhere, at any time with the aid of NVDs on virtually all its platforms. Understandably then, advances in NVD technology are crucial to widening the scope of night missions which in turn will keep U.S. forces 'owning the night.'

---

[1] A class A mishap is categorized by a loss of life or in excess of one million dollars property or casualty damage or both.

NVDs have generally been developed as single-band sensors, therefore constraining the user to the advantages and disadvantages of that band. However, researchers in the field of electro-optics have long known that gathering and melding information from two distinct EM bands (*sensor fusion*) would provide complimentary information to a user. (SPIE, 1987) They also knew that the process of sensor fusion would be computationally complex and therefore limited by the computer hardware required. In the late 1970's, British scientists and engineers seeking improvements over single-band NVDs suggested increasing advantages to pilots by fusing information from the I^2 and FLIR bands, combining it with a moving map and displaying it all on one wide field of view Heads-Up Display (HUD). This program, called "Nightbird," produced a flying fixed-wing platform which successfully demonstrated sensor fusion in aviation. (OPTEVFOR, 1993)

After "Nightbird," research with a comparable fusion system continued with a USN/USMC program called "Cheapnight." The results of this study proved the feasibility of using passive sensors to give fixed-wing platforms night attack capability but it did not specifically capitalize on the merits of sensor fusion. Follow-on studies like "Quicknight," "Fleetnight" and "Realnight" resulted in equipping formerly FLIR-only platforms with improved navigation FLIRs (NAVFLIR) and NVGs (e.g., A-6E Night Vision Imaging System). (OPTEVFOR, 1993) Correspondingly, formerly NVG-only platforms (mostly rotary-wing) are also being equipped with NAVFLIR (e.g., CH-53E Helicopter Night Vision System).

Recently, research in sensor fusion NVD displays has been rekindled. The general

aim of this research is to provide the best possible visual information to NVD users on a single display, thereby increasing capabilities while decreasing the workload of interpreting information from two or more displays. For example, the U.S. Army and Texas Instruments have in their inventory a rotary-wing platform equipped to provide the pilots with fused output from an image intensified charged-coupled device (I^2CCD) and FLIR. Additionally, a proposed Advanced Technology Demonstration (ATD), Color Night Vision System focuses on the additional benefits of artificially coloring the fused monochrome display (currently shades of phosphorous green) to possibly increase contrast cues in the output. The researchers hypothesized that this fused or fused color imagery will increase a user's situational awareness and therefore margin of safety and mission success. (Krebs, 1994; Scribner, et al., 1996)

Sensor fusion displays and their effectiveness in enhancing the night capabilities of military aircraft over current systems require detailed exploration in the areas of human factors and the mechanics of digital image fusion and enhancement.

This thesis is focused on the human factors of sensor fusion; more specifically, human perception of the fused and colored displays versus the IR and I^2 displays currently employed . The goal of this thesis is to quantitatively assess the impact of fused imagery and fused color imagery on human visual performance. Although one may gain an intuitive feel for image improvement simply by viewing NVD images before and after fusion or coloring, such intuitions are not quantifiable or adequately precise. Two precise measures of visual ability which are critical to aviation and the military in general are reaction time and

accuracy in target detection. This thesis developed a visual search experiment designed to employ static images from the four sensors involved (I^2, FLIR, fused monochrome and fused color) in measuring the impact on these variables.

Before offering an in-depth discussion of the experimental design or a quantatative assessment of the experimental output, there must be a structured presentation of the factors involved in constructing an NVD image as well as some of the physiological and psychological factors of human vision. Combining ideas from the "Sequence of events in the thermal imaging process" from Lloyd (1975) and the "Conditions for target acquisition" from the U.S. Army's NVEOL sensor model (MORS, 1995), a logical structure has been derived. The presentation follows the electromagnetic (EM) energy as it emanates from its source, through the atmosphere, through the sensor and ultimately how the output is perceived by the human observer.

B. NVD FACTORS

1. NVD Electromagnetic Spectrum

"The EM spectrum extends from barely measurable cosmic rays to electrical oscillations kilometers long. Electromagnetic radiation such as light, heat, x-rays, microwaves and radio waves are the parts of the spectrum humans depend on in their daily lives." (Lloyd, 1975) For the most part, the natural light from sunrise to sunset delineates a human's periods of activity and inactivity because the visual system cannot fully function outside the narrow band of visible EM radiation. NVDs, by processing EM bands not used by the human eye, enable exploitation of the night environment by the NVD user. These

devices do not turn night into day as will be shown later, but they do enable humans to better perform tasks as simple as night movement on foot or as complex as night attack in a high performance aircraft.

Current NVD images are processed from two distinct bands of EM radiation. NVGs process the visible and near IR spectrum (roughly 600 to 900 nanometers (nm)) and, much like the human eye, depend almost entirely on reflected energy for scene illumination. FLIRs generally process emissions from two infrared bands, midwave (3-5μm) and long wave (8-12μm). It is important to note that most man-made objects emit in the 8-12 μm band, hence the military interest in LWIR sensors. Figure 1 graphically illustrates the relationship of the EM bands used by NVGs, FLIRs (long wave IR shown) and the unaided human eye. The spectral bands are not where the differences end however, as the EM energy used by each NVD comes from differing sources and it is impacted by several variables en route to the receiving end of the particular sensor.

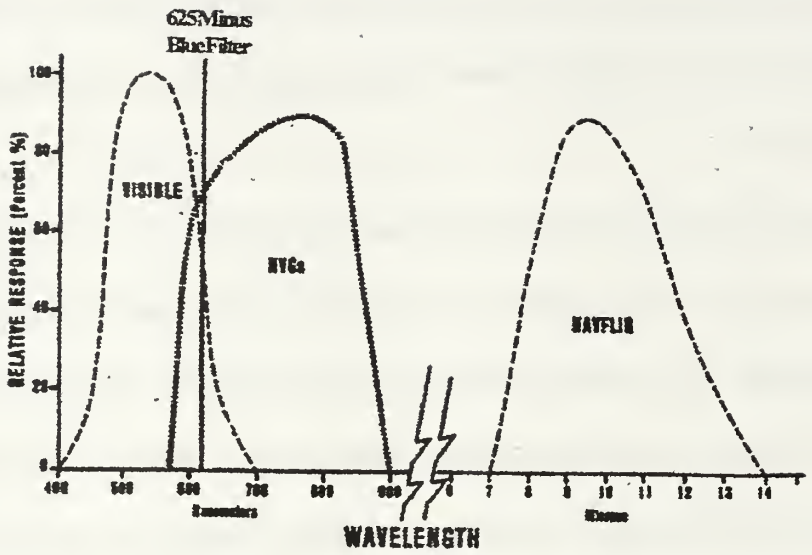

Figure 1. The Portions of the Electromagnetic Spectrum Used in Unaided Human Vision, by NVGs and by FLIRs. (MAWTS-1, 1995)

a. Optical Radiation

"Optical radiation (light), which is processed by NVGs, manifests itself in two ways; as particles of energy called photons or as waves. The particle theory of light provides a description of the emission of light from a source, such as the moon. The amount of light generated from a source (illuminance) is expressed in lumens per square meter or lux. The intensity of this energy, which is useful in dealing with the amount of reflected light available for NVGs, can be measured as the amount of light which strikes a surface. Reflected light (luminance) is expressed in terms of foot-lamberts (ftL). The wave theory of light is useful in describing the various phenomena having to do with the propagation of light through the air, or through an optical system such as the human eye. Regarded as a

form of wave motion, light has the characteristics of wavelength, frequency, and velocity." (MAWTS-1, 1995)

b. Infrared Radiation

"Infrared energy (thermal energy) is emitted by all objects with a temperature above absolute zero (-273 degrees Celsius). An increase in temperature will increase an objects molecular vibrational motion, thereby increasing its energy state. When the elevated energy state collapses, thermal energy in the form of radiation is emitted. In general, thermal radiation which strikes an object can be absorbed, transmitted or reflected. Natural thermal energy is produced when objects absorb thermal energy from IR sources such as the sun or warm air currents. Once absorbed this energy can then be radiated. Another source of thermal energy is from man-made objects such as the heat radiated from a running engine or the heat radiated as a result of the friction from moving parts." (MAWTS-1, 1995) IR radiation is independent of optical radiation but is more complex and requires additional discussion on one of the most important factors impacting an object's temperature, its 'emissivity' (E).

In order to comprehend emissivity, one must have a standard from which to start. In thermodynamics that standard is called a 'blackbody.' "Blackbodies are defined as the perfect absorber of thermal energy and are therefore also perfect emitters, with an efficiency of unity." (MAWTS-1, 1995) Emissivity then is the ratio of an object's ability to emit thermal energy at a certain temperature over that of a blackbody at the same temperature. Other factors impacting emissivity are: Material composition, surface finish,

ambient temperature and the object's temperature and geometry. Most natural objects have a high emissivity and therefore a majority of their thermal signature is from self-emission. Emissivities of some common materials are listed in Table 1. Conversely, objects with low emissivity have a corresponding high reflectivity and therefore reflect thermal energy of their surroundings.

MATERIAL	EMISSITIVITY
HIGHLY POLISHED SILVER	0.02
HIGHLY POLISHED ALUMINUM	0.08
POLISHED COPPER	0.15
ALUMINUM PAINT	0.55
POLISHED BRASS	0.60
OXIDIZED STEEL	0.70
BRONZE PAINT	0.80
GYPSUM	0.90
ROUGH RED BRICK	0.93
WHITE LACQUER	0.95
GREEN OR GREY PAINT	0.95
LAMP BLACK	0.95
WATER	0.96

Table 1. Emissivities of Some Common Materials. (MAWTS-1, 1994)

The discussion presented up to this point has focused on delineating the spectra used by NVGs and FLIRs and the theories of optical and infrared radiation. The following section will focus on energy sources and the energy as it moves toward the NVD.

2. NVD Scene Variables

Mission planning considerations for the use of NVDs far exceed those for daylight missions. The first and foremost planning consideration is the quantity and quality of a

specific EM bandwidth that can be expected as this is the basis for the NVD scene that will be displayed. Planners must consider the energy's source, any media the energy may pass through, any attenuation that may occur and any objects the energy may impact as it travels to the sensor. Accordingly, these planning considerations can be grouped into three main categories: (1) sources, (2) terrain effects and (3) atmospheric effects. The following subsections will discuss these considerations for optical and infrared radiation as they apply.

a. Sources

(1) Optical radiation. Illumination, measured in lumens per square meter (lm/m^2) or lux, is one of the sources of energy that NVG's intensify; however, it has no impact on FLIR imagery. The moon provides a reflection of seven percent of the sunlight that strikes it, making it the largest and brightest natural object in the night sky when it is visible. Lunar illumination then is the primary energy source for natural illumination in the night sky (Figure 2). (MAWTS-1, 1995) Another significant contributor to nighttime illumination is the moonless night sky with various stellar phenomena. Additionally, starlight contributes up to .00022 lux (1/10th the level of a quarter moon) while auroras, zodiac lights and other phenomena of the atmosphere provide even smaller contributions. Figure 3 illustrates how moonless night sky illumination almost matches the peak sensitivity of NVGs.

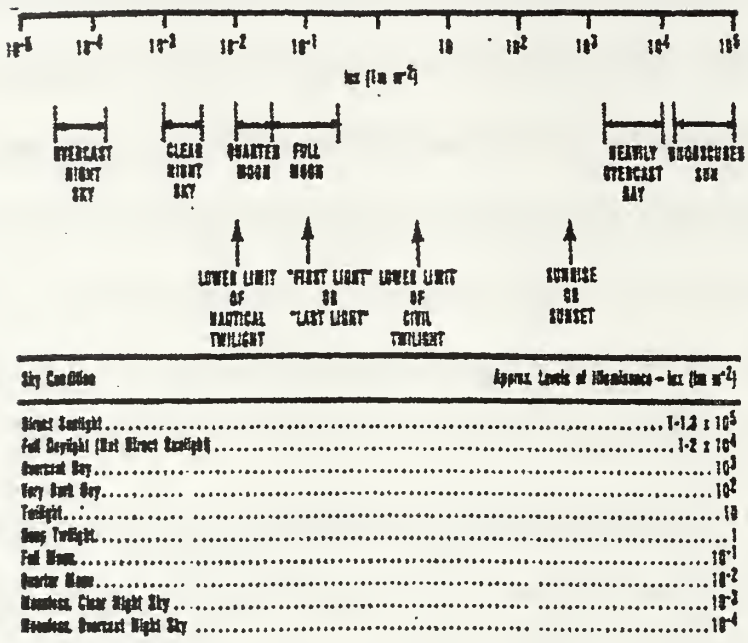

Figure 2. Illumination Levels of the Moon and Sun. Lux Levels Contributed by Each Source Are Listed. (MAWTS-1, 1995)

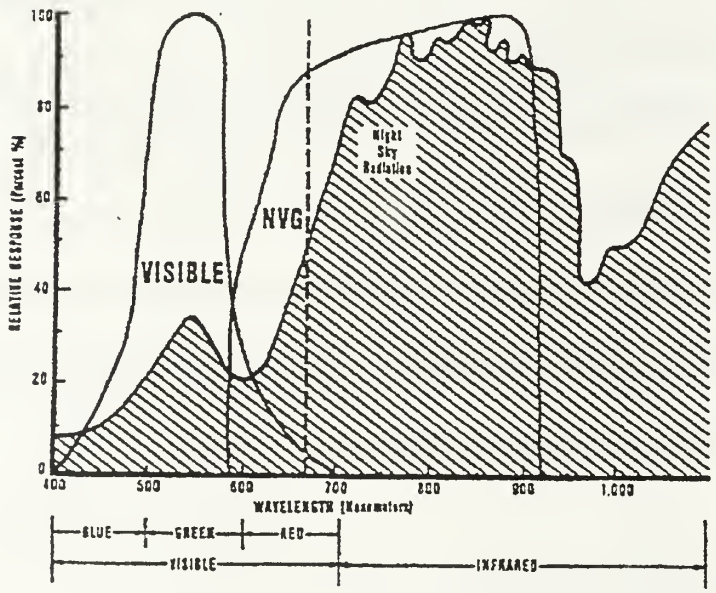

Figure 3. Night sky illumination overlaid with NVG and unaided vision peak sensitivities. (MAWTS-1, 1995)

Two other contributors of illumination that may be more of a hindrance than a help are the sun and artificial (cultural) sources. The setting sun at zero to six degrees below the horizon is too bright for NVG operations, however, approximately one half hour after sunset, when the sun has lowered to seven degrees below the horizon, it may provide useable illumination. This useable illumination period continues until the sun has set past twelve degrees. Artificial lighting such as street lights or radio tower warning lights can also provide significant illumination, however, cultural areas with large concentrations of artificial illuminators can wash out the NVG image. Illumination impact on NVG output will be discussed further in the section on Contrast Sensitivity.

(2) Thermal radiation. Thermal energy sensed by FLIRs is measured in microns (μm) and is invisible to NVGs. The three principle sources of thermal energy mentioned earlier are solar radiation, fuel combustion and frictional heat, and thermal reflection. Solar radiation is one of the most prominent contributors to the thermal signature of objects exposed to it. Given that an object is exposed to the sun on a clear day, then the location on the earth, the time of day and the time of year will determine the intensity of solar radiation. Fuel combustion and frictional heat sources generally emit a higher thermal signature than their surroundings. These blooms of thermal energy or 'hot-spots' exceed the boundaries of the source and, in that respect, their signature overtakes nearby emissions of lesser value. (MAWTS-1, 1995) The impact of hot-spots on the output of IR sensors will be discussed further in the next subsection.

The last source of thermal energy is that which is reflected. Objects

with low thermal emissivity possess a corresponding high thermal reflectivity. In the night environment, a horizontally oriented object of low emissivity (e.g., a body of water) will reflect the thermal energy of the night atmosphere above it and appear cooler than its surroundings. Conversely, a vertically oriented object of low emissivity (e.g., a canyon wall) will reflect the temperature of its surroundings and therefore blend into the thermal scene. (MAWTS-1, 1995)

The discussion thus far has focused on the primary sources of optical and thermal radiation. Regardless of the source or the sensor, the main concern of an NVD user is how the device will improve functions like navigation, terrain avoidance and target detection. Accordingly, the next section will delineate the effects on EM energy as it is reflected or radiated by the collection of objects on the earth's surface which, for simplicity, will be called 'terrain.'

b. Terrain Effects

Optical radiation leaves a source as photons and propagates until it impacts objects in its path. The ratio of the light that is reflected by an object over the amount of light that is incident to it is called its albedo or reflectivity. Reflected light or luminance is measured in foot-Lamberts (ftL). Table 2 lists the albedos of some common terrain.

SOILS	DRY	WET	WET/DRY
Dark	0.13	0.08	
Light	0.18	0.10	
Dark - plowed	0.08	0.06	
Light - plowed	0.16	0.08	
Clay	0.23	0.16	
Sandy	0.25	0.18	
Sand	0.40	0.20	
White sand	0.55		
SURFACES	**DRY**	**WET**	**WET/DRY**
Asphalt			0.10
Lava			0.10
Tundra			0.20
Concrete			0.30
Stone			0.30
Desert			0.30
Rock	0.35	0.20	
Dirt Road	0.25	0.18	
Clay Road	0.30	0.20	
FIELDS	**GROWING**	**DORMANT**	**EITHER**
Tall Grass	0.18	0.13	0.16
Mowed Grass	0.26	0.19	0.22
Desiduous Trees	0.18	0.12	0.15
Coniferous Trees	0.14	0.12	0.13
Rice	0.12		
Best Wheat	0.18		
Potato	0.19		
Rye	0.20		
Cotton	0.21		
Lettuce	0.22		
SNOW			
Fresh	0.85		
Dense	0.75		
Moist	0.65		
Old	0.55		
Melting	0.35		
ICE			
White	0.75		
Gray	0.60		
Snow & Ice	0.65		
Dark Glass	0.10		

Table 2. Albedos of Some Common Terrain (ftL). (OPTEVFOR, 1993)

Reflectivity plays an important part in what is visible in the optical radiation spectrum and what is not. Two different terrain surfaces may have two different reflectivities and therefore exhibit terrain contrast. Another factor in terrain contrast is the texture of the terrain. Because of texture, terrain that has considerably low reflectivity can provide recognition and depth perception cues over that available from terrain with higher reflectivity (e.g., forest over desert). Understandably, the less illumination available, the less terrain contrast and visual scene that can be expected. Terrain blocking illumination from other terrain is where there would be no reflection and therefore no terrain contrast. As in the day environment, this is called shadowing but it is more significant at night because shadowed objects can be effectively hidden from view. A dangerous example of shadowing in the NVG environment is the aircraft flying toward what appears to be a tall mountain being highlighted by the low-angle moon. Lurking in the shadows, however, is the shorter but closer mountain that presents an impact hazard.

Thermal energy is either emitted or reflected by an object but it is primarily the temperature difference of objects that make up the thermal scene. If there is no temperature difference between objects on the terrain, then the terrain appears homogeneous in the thermal scene. This is not usually the case as the sun provides solar radiation to the terrain in the daylight hours and none at night. The cyclic heating and cooling of the terrain causes the diurnal cycle of temperature differences between objects of different thermal mass and inertia.

Figure 4 shows the diurnal cycle of temperature differences for an armored

vehicle and other objects considered as background terrain. From the graph one can visualize the negative thermal contrast (object cooler than background) of the armored vehicle on a clear sunny day and the positive thermal contrast (object warmer than background) of the armored vehicle at night. Crossover times, when the temperature of the

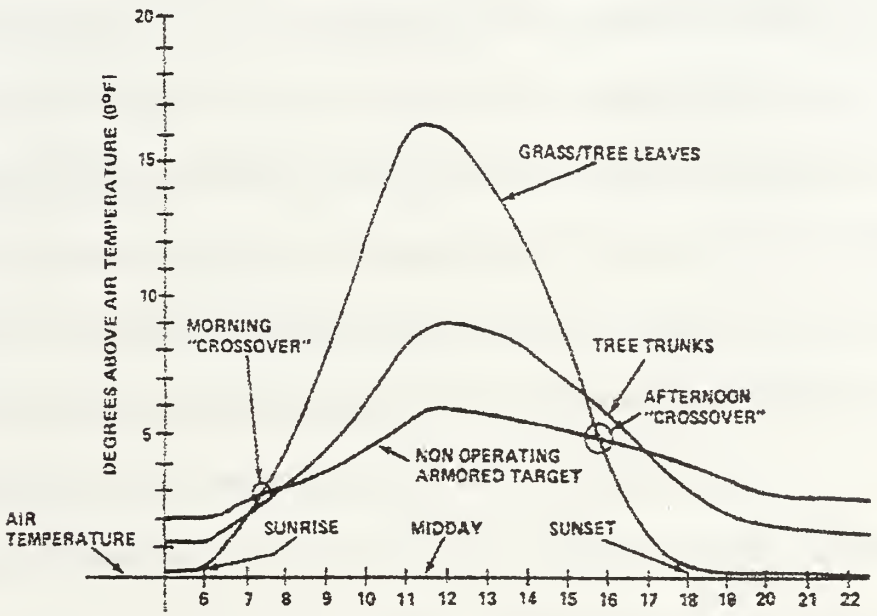

Figure 4. A sample diurnal cycle for a man-made object and the background terrain. Crossover times shown. (MAWTS-1, 1994)

object equals that of the background are depicted. Even on overcast days, some solar radiation is absorbed by the terrain which in turn continues the diurnal cycle.

Another small contributor to the thermal scene is thermal shadows. Thermal shadows are present as the terrain cools at sunset but they dissipate quickly. Thermal energy from combustion and friction is usually hotter and more persistent in the man-made object

than solar radiation. When the object moves, the thermal footprint of where it has been is left behind and is detectable sometimes for hours. This footprint may cause a thermal decoy for someone trying to detect an object using a FLIR.

Radiated or reflected energy, after it leaves its source, must travel through a medium en route to a sensor or an intermediate object; the medium is the earth's atmosphere and the next section will be a discussion of its impact.

c. Atmospheric Effects

The most significant impact on the optical and thermal energy available for NVDs is made by the atmosphere. In the atmosphere, attenuation of energy after it leaves the source can occur by refraction, absorption or scattering. Because attenuation by refraction is negligible, only attenuation by absorption and scattering will be discussed.

(1) Absorption. Attenuation by absorption is more significant than that by scattering. Absorption of EM energy for NVDs centers around three atmospheric molecules; water, carbon dioxide and ozone. Of the three molecules, atmospheric water vapor or humidity is the most significant absorber. In very hot and humid climates, the high amount of absorption may literally render the FLIR useless. (MAWTS-1, 1995) Carbon dioxide is second to water in absorbing capability but it is usually in a uniform concentration in the atmosphere. This uniform concentration makes predicting its impact much easier than the erratic effects the lowest absorber, ozone. Ozone only absorbs thermal energy and its natural influx from the upper atmosphere is extremely difficult to predict. Man-made sources such as industrial pollution or combustion products are sources of ozone that may

be predicted when flying near industrial or dense urban areas. (MAWTS-1, 1995)

(2) Scattering. Light and heat traveling through the atmosphere can impact objects or molecules and be scattered in different directions. There are two types of scattering; molecular and aerosol. Molecular scattering occurs when light strikes particles that are smaller in wavelength than the light itself. Nitrogen, oxygen, water vapor and carbon dioxide all meet this requirement. (MAWTS-1, 1995) Aerosol scattering takes effect with particles larger than one micron, such as dust, smog, snow and other natural or man-made obscurants. Because of its longer wavelength, thermal radiation is not significantly impacted by aerosol scattering. (MAWTS-1, 1995)

After considering optical and thermal energy sources, the energy that is emitted and what impacts the energy as it propagates through the atmosphere, the night vision devices that sense this energy may be discussed.

3. The Sensors

a. NVGs

"NVGs are electro-optical devices used to detect and intensify optical images in the visible and near infrared region of the EM spectrum for the purpose of providing visible images." (MAWTS-1, 1995) Current NVG technology centers around the third generation (Gen III) image intensifier (I^2) tube. Although the electronics of image intensifiers is beyond the scope of this thesis, a basic explanation of the functions of the five major components of an I^2 device and how they turn optical energy into useable output is necessary. Figure 5 shows three of the five major I^2 components; the photo cathode, the

microchannel plate and the phosphor screen. Not depicted in Figure 5 are the objective lens on the front of the tube and the eyepiece lens on the back.

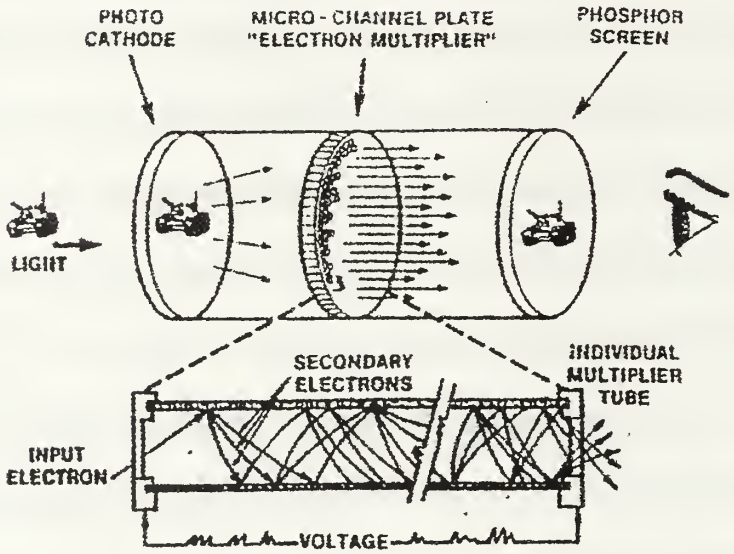

Figure 5. A Basic Image Intensifier (Objective Lens and Eyepiece Not Shown). (MAWTS-1, 1994)

Radiant or reflected optical energy first strikes the objective lens of the I^2 tube where it is focused onto the photo cathode. The photo cathode, which is made up of gallium arsenide crystals, detects optical energy in the near IR to visible spectrum (600-900 nm) and converts this energy into electrons. Electrons accelerating forward from the photo cathode strike the 'intensifier' part of the tube, the microchannel plate. The microchannel plate increases the number of electrons at its output by a factor of one thousand. Electrons entering the front of millions of specially lined microscopic glass tubes that make up the

plate are deflected numerous times as they travel the length of the tubes, causing secondary electron emissions. The resultant electrons accelerate toward the phosphor screen from their respective tubes, maintaining their relative spatial position. (MAWTS-1, 1995)

The phosphor screen consists of a thin coating of phosphor on the input end of a wafer-thin fiber optic image inverter. The phosphor screen turns the electrons impacting it into yellow-green light in the 560 nm range, matching the peak sensitivity of photopic human vision. The image inverter takes this light and inverts it by way of a 180 degree twist in the fibers. The image inverter also serves to collimate or focus on infinity the image being sent to the eyepiece lens; without this the user's focus would be on the eyepiece lens, causing severe eye strain. (MAWTS-1, 1995)

The final component of an I^2 device is the eyepiece lens. The eyepiece lens serves to focus the output image from the phosphor screen onto the human eye by way of an adjustable diopter ring. The ratio of the brightness of the image at the output of the eyepiece lens over the luminance of the light entering the objective lens is called the 'gain' of the I^2 device. The variants of NVGs depicted in Figures 6 and 7 employ Gen III I^2 tubes with a gain of 25,000, a substantial advantage over the unaided human eye in the night environment. (MAWTS-1, 1995)

Figure 6. A Fixed-wing Aviator Equipped With MXU-810/U Cats Eyes NVGs. (MAWTS-1, 1994)

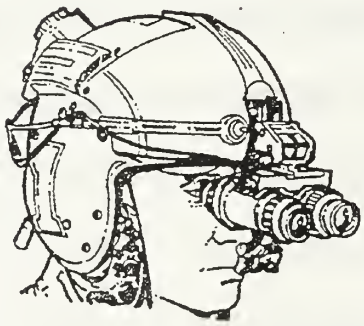

Figure 7. A Rotary Wing Aviator Equipped With the AN/AVS-6 Aviators Night Vision Imaging System (ANVIS). (MAWTS-1, 1995)

b. FLIRs

FLIRs are electronic devices that convert invisible energy from the far infrared spectrum into a visible image. All FLIRs are temperature differential sensors that are adjusted to sense a range of temperatures called the sensor's 'gain.' Military FLIRs allow a gain as wide as 90 degrees Celsius. An important measure of performance of a FLIR is 'delta T' or the temperature difference of an object and its background. A FLIR's gain setting determines thermal sensitivity and the delta T. (MAWTS-1, 1994)

Current FLIR technology is centered around the first generation (Gen I) FLIR thermal imaging device. FLIR systems are complex and varied and their electronics are beyond the scope of this thesis; however, a basic explanation of the functions of the three major components of a thermal imager and how they convert thermal energy into useable output is necessary. Navigation FLIRs (NAVFLIRs), which will be discussed here are different from other FLIRs in that they provide the user with a thermal scene the size of the NAVFLIR field of view. Figure 8 shows the three major NAVFLIR components; the infrared sensor, the the signal processor and the cockpit display. (MAWTS-1, 1994)

The infrared sensor has many important subsections that are critical to gathering thermal energy. First, an IR window must be present to protect the sensor while allowing the 8-12 μm EM energy to pass through to the IR telescope. Germanium or other IR transmissive materials are used for this window. The IR telescope functions to focus a thermal scene comprable in size to the field of view of the cockpit display onto the motor driven scan assembly. (MAWTS-1, 1994)

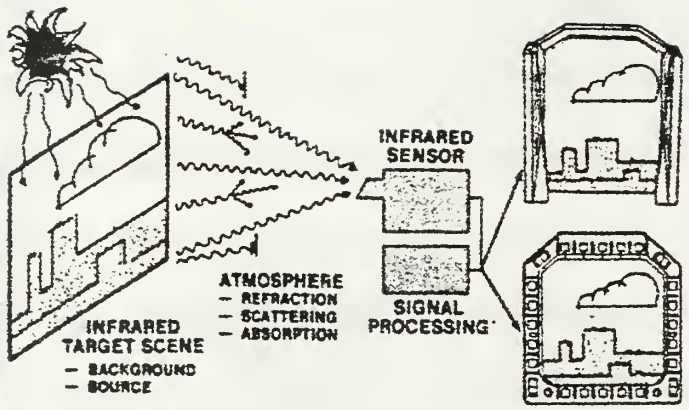

Figure 8. A basic NAVFLIR (Heads-up Display, upper right; Heads down display, lower right). (MAWTS-1, 1994)

A scan (mirror) assembly serves to rapidly transfer the thermal scene provided by the IR telescope onto a photoconductive detector array. NAVFLIR detector arrays are quantum detectors tuned to sense as little as one degree celsius delta T. In order to provide this thermal sensitivity, the array is continuously cryogenically cooled. The detector array is composed of semiconductive material which turns 8-12 μm heat energy into analog electrical output to the signal processor. Each detector in the array has its own channel for analog output. (MAWTS-1, 1994)

The signal processor, depending on the model NAVFLIR, performs many varied functions but basically it provides the special signal functions required to stabilize and enhance the analog output from the detector array so that it is suitable for display in the cockpit. The signal from the signal processor is transformed to an image through the use of a cathode ray tube (CRT) and the color of the image is a function of the phosphor used in the CRT. Cockpit displays can be either a heads down display (HDD) employing a CRT,

a heads up display (HUD) employing a combiner glass to provide a see-through reflection of the CRT image or a helmet mounted display (HMD) employing a mini-CRT on a monocular assembly. (MAWTS-1, 1994) Figures 9 and 10 are examples of current military NAVFLIR systems.

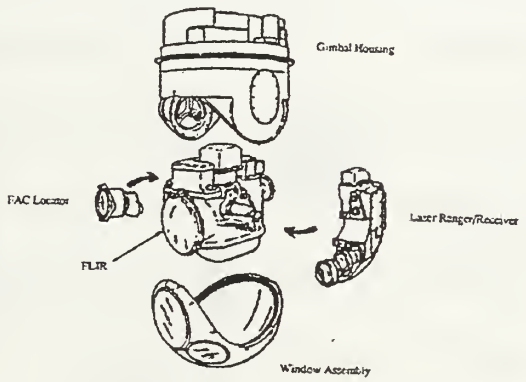

Figure 9. The A-6E Detection and Ranging Set Employing a Gen I NAVFLIR. (OPTEVFOR, 1993)

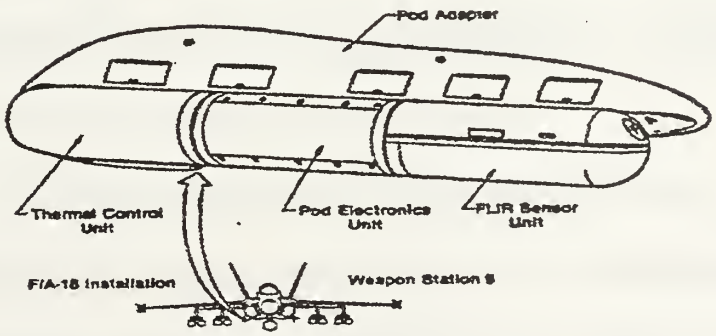

Figure 10. The F/A-18C/D AN/AAR-50 Gen I NAVFLIR. (MAWTS-1, 1995)

c. Fused Monochrome and Fused Color

I^2 and FLIR sensors provide complimentary visual information that enhances human effectiveness during night operations (Figure 11). It is hypothesized that combining the images from these two sensor bands to provide a single fused display will significantly improve performance using NVDs above current capabilities.

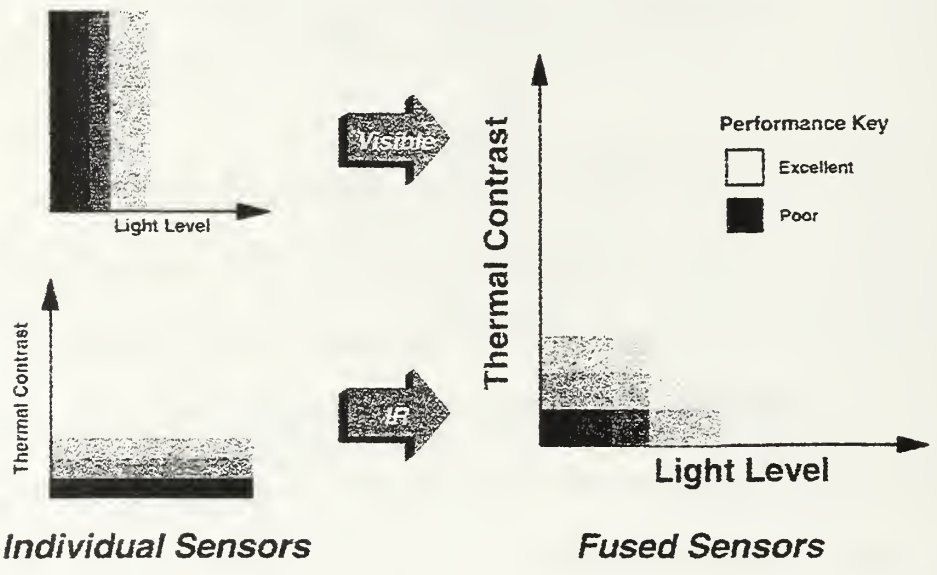

Figure 11. The Complimentary Nature of IR and I^2 (Visible) Information. (Courtesy of NVSED)

(1) Fused Monochrome. The improved performance with sensor fusion is based on the rattlesnake visual system which combines visible and infrared vision for hunting at night with little or no light. The snake's visual system is composed of infrared sensors, pit organs, located near the head that open on the side of the head below and in front of the eyes. Infrared information is sensed by the pit organs and is then sent to the brain

where it is combined with visible information obtained from the snake's eyes. All snakes of the subfamily Crotallinae, pit vipers, have pit organs that are sensitive to infrared information. (Newman & Hartline, 1982) Laboratory experiments have shown that the pit viper could distinguish between a warm light bulb covered with an opaque cloth and a cold bulb. The snakes struck the warm bulb as long as their pit organs were not obstructed, if the organs were covered then the snake ignored both the warm and cold bulb (Noble and Schmidt cited in Newman & Hartline, 1982). This same integration of visible and infrared information in the pit viper may also prove useful for military forces operating at night.

The Army Night Vision Electronic Sensors Directorate (NVSED) and Texas Instruments (TI) proposed a sensor fusion system that would combine an I^2 sensor and a Gen I FLIR sensor within a UH-1N aircraft to enhance helicopter navigation. (Texas Instruments and U.S. Army, 1993) This Advanced Helicopter Pilotage System (AHPS) is presently mounted on a UH-1N helicopter and has provided some of the imagery used in this thesis. NVSED and TI hypothesized that the AHPS would combine the optimal information from the two sensor spectral bands and would therefore increase visual performance as supported by the pit viper's enhanced night vision model.

One of many fusion techniques available is the modified Peli-Lim algorithm, which basically separates the high and low pass image components, boosts the low-pass (low luminace value pixels) portion and then recombines it with the high-pass components (Figure 12). The resultant signal is relinearized to an 8-bit fused monochrome image. Like the I^2 and FLIR sensors, integral to a fusion device, the electronics involved

Peli/Lim Fusion Algorithm

(Monochrome)

Adaptive Filtering for Image Enhancement

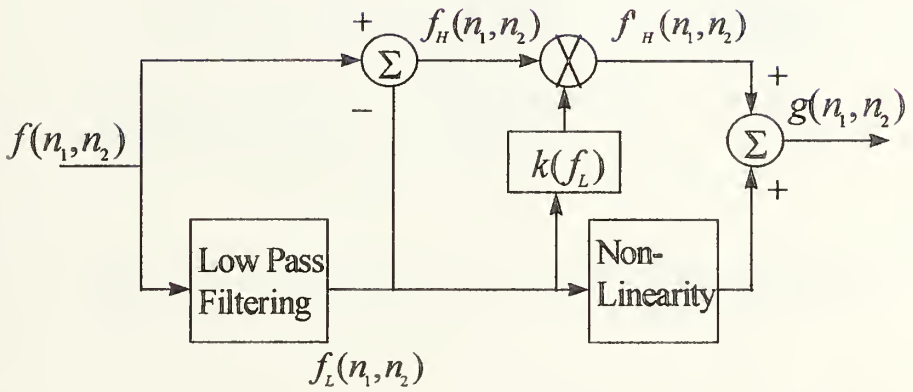

Figure 12. Peli/Lim Fusion Algorithm. High and low pass elements of an image are separated. The low pass element is boosted and recombined with the high pass element. The recombined output is relinerized to an 8-bit image. (Courtesy of CVSAD)

with fusing the outputs is beyond the scope of this thesis. However, the three major components of the existing Army/TI device and their functions will be discussed. Figure 13 depicts the three major components of the AHPS; the sensors, the fusion processor and the cockpit display.

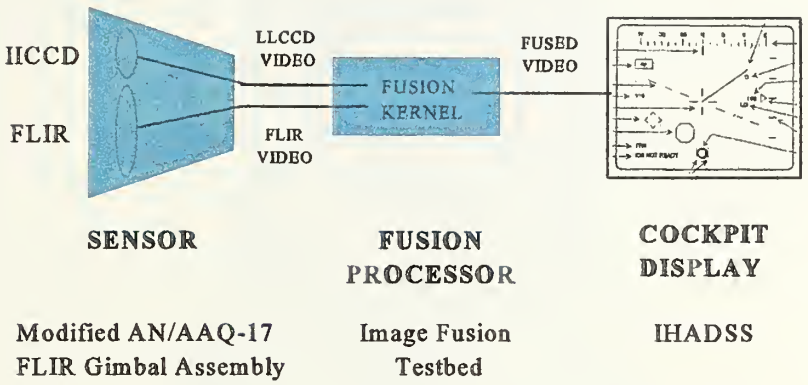

Figure 13. Schematic of The Advanced Helicopter Pilotage System (AHPS). The Three Main Components Shown.

A concept similar to the AHPS sensor head is shown in Figure 14. The sensor head of a modified Lockheed-Martin "Nitehawk" IR pod is equipped with an image intensified charged coupled device (I^2CCD) integrated into the gimbal assembly. The I^2CCD gets its name from the Gen III I^2 tube whose luminous output is fed through a fiber coupling to the TV sensor, producing the I^2TV video output. The FLIR uses standard FLIR technology to provide FLIR video output. In the AHPS, the two video outputs are fed into the fusion processor where the individual video inputs are preprocessed and optimized by weighting each sensor's localized pixel array depending upon a weighting criteria. For example, if the registered I^2CCD image appeared to be better than the registered IR image, then the fusion device would receive 60% input from the I^2CCD pixel and 40% input from the FLIR pixel. (Texas Instruments & U.S. Army, 1993)

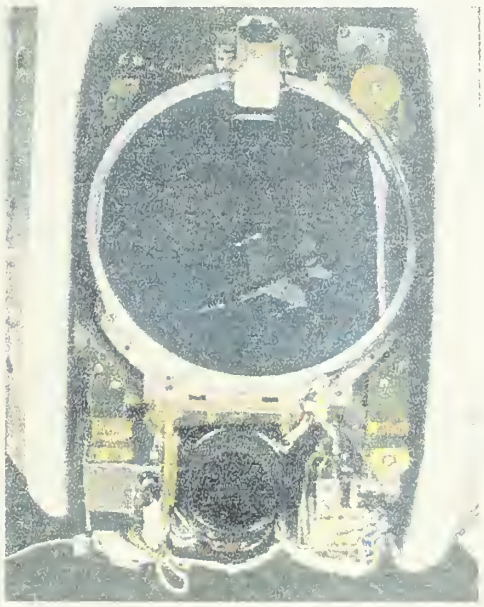

Figure 14. A modified "Nitehawk" FLIR gimbal assembly with an integrated I^2CCD (lower lens). (Courtesy of Lockheed-Martin)

The resultant optimal fused video is provided to the aviator in the cockpit through a modified Integrated Helmet and Display Sighting System (IHADSS). The IHADSS provides a monocular output to the pilot corresponding to where the pilot is looking. Because there is only one AHPS assembly, only one pilot at a time can control the IHADSS with their head motion.

(2) Fused color. Krebs (1994) and the Naval Research Laboratory (1995) proposed an extension of NVESD and TI's program by providing an alternative processing technique that would display a color scene instead of a monochrome greyscale image. They hypothesized that using concepts of human biological vision ('opponent' color cells or cells that sense colors that do not naturally mix), color contrast cues would allow

separations of vegetation, sky, water, ground, and the identification of targets in various lighting by terrain. Figures 15-17 are diagrams with amplification provided as a tutorial by NRL to enable a clear understanding of the otherwise complicated color fusion process.

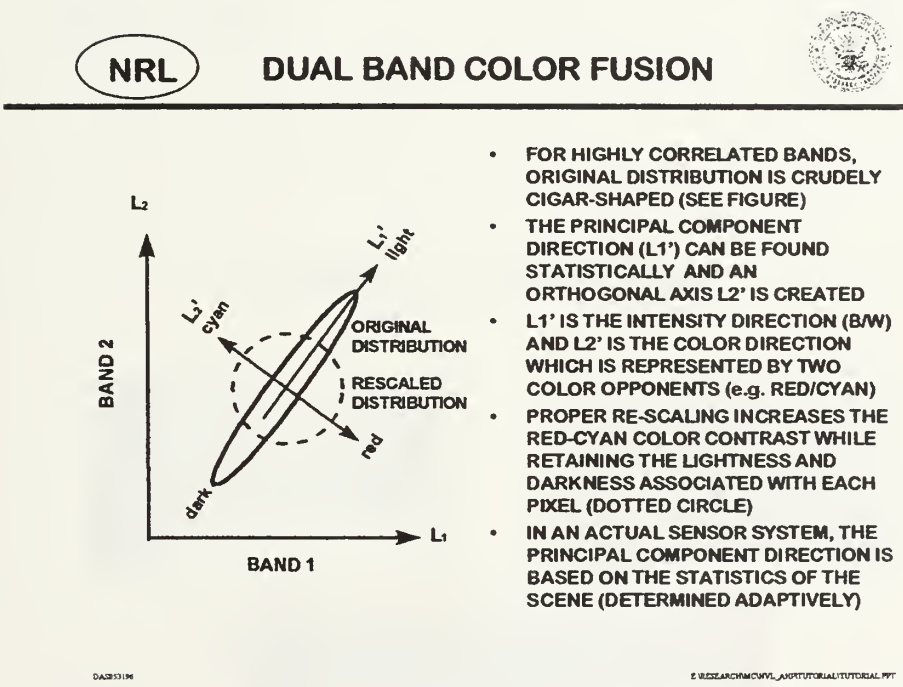

Figure 15. Dual Band Color Fusion Diagram With Amplification. (Courtesy of NRL)

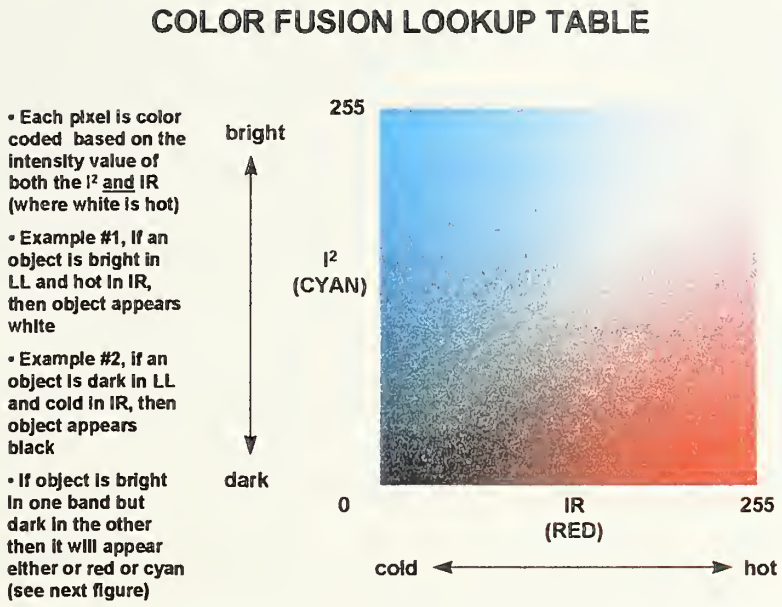

Figure 16. Color Fusion Look Up Table (LUT) With Amplification. (Courtesy of NRL)

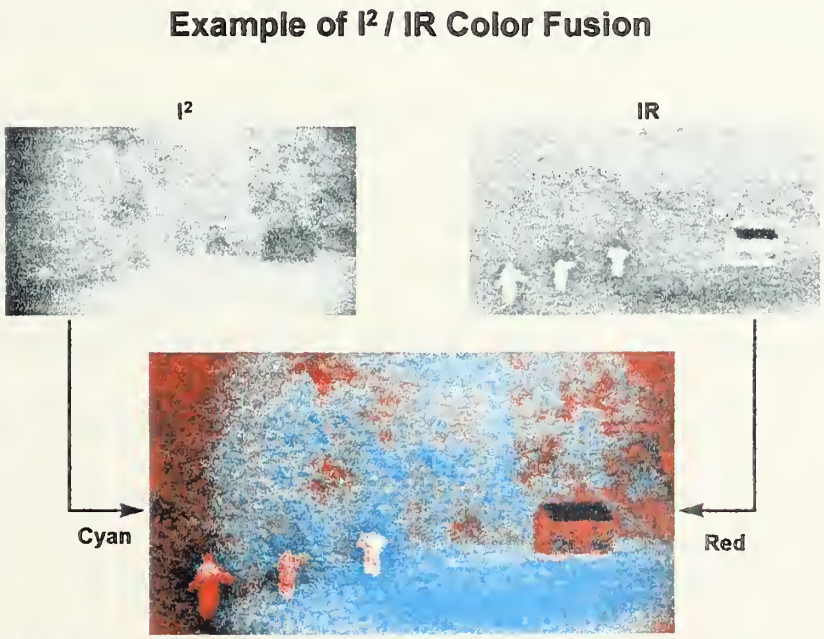

Figure 17. An Example of Color Fusion. (Courtesy of NRL)

With a fused color system, the three major components would theoretically remain the same as the fused monochrome system (See AHPS, Figure 13) except that the fusion and coloring would be done in a combined 'color fusion' processor. Also the IHADSS or other display would require a color CRT modification for the output.

This section has focused on the energy that NVDs sense, on the physical sensors themselves and how they produce their output image. More information on how the output imagery is generated and discussion on the merits of each type of output is required and is presented in the next section. Also, understanding the impact these sensors have on visual performance is imperative to measuring improvements from one sensor to another. Accordingly, the next section will cover the human factors of using these devices.

C. HUMAN FACTORS

1. Situational Awareness

The main effect of wearing NVD's for aviators and others is the increased situational awareness over night unaided flight. Situational awareness is defined in the MAWTS-1 Helicopter Night Vision Device Manual as "the degree of perceptual accuracy achieved in the comprehension of all factors affecting an aircraft and crew at a given time." (MAWTS-1, 1995) During daylight flying with few visual obstructions, pilots have many visual cues available to them, however, these cues are ones for which their photopic (day) vision was optimized and they are quickly used to improve the pilot's situational awareness. "The first consideration that must be emphasized with NVDs is that they do not allow you to assume a daylight posture for mission planning or execution. NVDs should be treated as a very

reliable, very accurate instrument, but as with all other instruments, it must be continually crosschecked with other instruments and or crewmembers to get an accurate assessment of the real world." (MAWTS-1, 1995)

Since the greatest aeromedical concern of NVD operations is the effect these devices and the night environment have on the human visual system, one must have a basic understanding of this system and how visual cognition is used to keep humans situationally aware.

2. Visual Cognition

a. Parallel Processes

> *"Whatever we know about reality has been mediated, not only by the organs of sense but by complex systems which interpret and reinterpret sensory information."* Ullrich Neisser, 1967

Ullrich Neisser (1967) demonstrated that humans have the ability to store visual input in some medium (iconic memory) which is subject to rapid decay. Before it has decayed, information can be read from this medium just as if the stimulus were still in view. He discovered empirically that iconic memory was found to be affected by visual variables like intensity, exposure time and post exposure illumination. Also, he found the useful life of the icon depended nonlinearly on exposure intensity and time (the useful life was not identical to exposure time) and that the duration of iconic memory was affected greatly by post exposure illumination.

With regard to human visual perception then, Neisser made the innovative

discovery that perception is an evolutionary and dynamic process. This discovery is still the accepted model in vision research and has been the basis for contiued studies on target detection and accuracy.

As previously stated, Neisser found that bright post exposure illumination significantly decreased human visual perception from the visual icon formed. He derived his findings from the results of visual tachistoscopic (t-scope) experiments coupled with a technique called "backward masking." T-scope experiments involve a subject veiwing a stimulus presented for a brief period of time. Backward masking involves flashing a "mask" (usually a black and white checkerboard) immediately after the stimulus in order to produce varying levels of degradation or erasure of the previous icon - the less similar the mask and icon, the more the degradation and vice versa. "Fortunately for humans, backward masking is not apparent in the everyday visual experience due to relatively small amounts of eye movements per second (i.e., five for reading) and long periods of fixation. Increasing eye movements to ten per second would make it impossible for humans to see anything well." (Neisser, 1967)

Neisser drew upon the results of backward masking experiments in which the subject had no indication of trials where the stimulus would be followed by a mask, but they were always required to respond quickly if they saw 'something.' The results of these experiments showed that rapid responses were no slower for the masked stimuli versus the unmasked stimuli, which meant that subjects had received enough visual information to respond in either case. On this finding Neisser wrote:

> "This rather dramatic result shows that visual information is processed in several different ways at once, "*in parallel*." While the construction of contours has only begun at one level, a message that "something has happened" is already on its way to determine a response. In this situation, the subject's response is not dependent on his having "seen" the stimulus figure clearly. It is only necessary that some sort of visual activity be initiated. This saves many milliseconds of response time with clear biological advantages." ... "Visual cognition is not a single and simple interiorization of the stimulus, but a complex of processes."(Neisser, 1967)

Neisser elaborates on the "complex of processes" in visual cognition and describes a 'wholistic (also holistic)' or 'preattentive' process where information in the human field of view is constantly being received and images are being constructed and synthesized in a hierarchical manner (i.e., motion then shape may be a possible heirarchy). By synthesis he meant that once a 'visual snapshot' is formed in the human brain, the information from it is incorporated into what the human sees rather than being retained as a separate entity. This is intuitive because vision as we know it would be impossible as a series of overlaid snapshots.

The visual demands of an aviator are complex and involve this holistic visual processing conducted at a rate much faster than with the relatively stationary human on the ground. Such visual abilities are further characterized in the more current vision research literature as "preattentive processing" tasks. (Triesman, 1985) Preattentive visual processing mediates human abilities that require rapid, parallel, assessment of the visual image. (Julesz, 1984; Treisman, 1985; Essock, 1992;) This preattentive image processing is required to segment the image into objects; into foreground and background, horizon and background, or target from background image structure, thereby establishing a rapid spatial

representation of the visual environment.

As developed by Julesz (1984) and others, low-level pixel, or pixel-cluster, information is used by the human visual system to characterize image regions, form meaningful regions, and possibly permit regions to 'pop-out' from the background with no conscious effort. (Essock, 1992) Neisser believed that as the construction and synthesis proceeds to a point which peaks human interest, the human will train their visual focus (fovea) on that form for additional processing and detailed recognition.

b. Experience and Prior Expectancy

Neisser hypothesized that a great deal of what does receive higher processing is recognized as a result of 'experience' and 'prior expectancy.' A simple example of experience is letter recognition which is done easily by literates but which is virtually impossible for illiterates because they have no basis for further synthesis and segregation of the forms on the paper. An example of prior expectancy would be expecting an "n" to follow "coi" and therefore form the word "coin." (Niesser, 1967)

In aviation, pilots are trained extensively on scene interpretation in a ground school and in actual flight. However, much like the literacy example above, the higher-trained users will be able to 'read' the scene below and navigate to an objective where the novice or person less trained would surely get disoriented. The higher-trained users will also know what to expect when they are correlating terrain represented on the map and what is before them on the ground. Prior expectancy also plays a major part in an NVD user's survival in the night environment due to the intense training regimen required for

survival in the night environment due to the intense training regimen required for interpreting the limited but dynamic information being viewed on the output device.

In summary, the work of Neisser and others since 1967 has contributed greatly to understanding human visual cognition and has provided fertile ground for studies in the higher level cognitive processes involved in target detection. Two studies in modeling early human vision, based on Neisser's work, focused on the substances of early vision and texture segmentation respectively. Because these studies are fundamental to understanding the methods and conclusions of this thesis, they will be discussed in the subsections below.

3. The Plenoptic Function

In their research on early visual processes, Adelson and Bergen (1991) noted the general concensus of researchers concerning the model of the first stage of human and machine vision (in the style of Niesser, 1967). Figure 18 illustrates the basic image properties or parallel pathways that comprise this model.

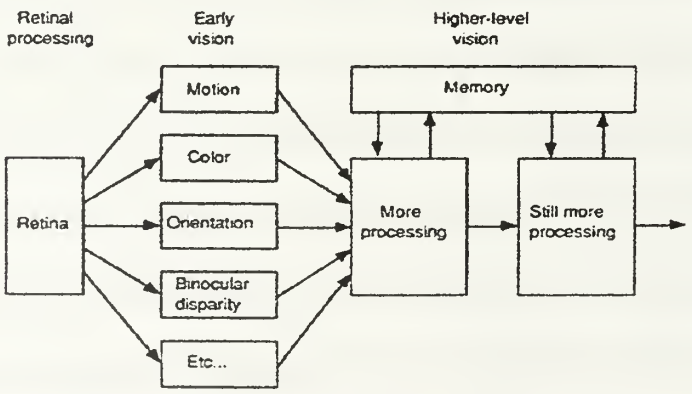

Figure 18. An Accepted Model Of Early Human Vision. (Adelson and Bergen, 1991)

In their research, Adelson and Bergen sought to take this accepted model of structured elements and break it down further to the substances of vision. "In other words, we are interested in how early vision measures 'stuff' rather than how it labels 'things'." (Adelson and Bergen, 1991) To accomplish this task, the authors formulated a function that would allow systematic derivation of the visual elements and provide a relationship of these elements to the strucure of visual information in the world. In describing the function, they wrote, "We will show that all the basic visual measurements can be considered to characterize local change along one or more dimensions of a single function that describes the structure of the information in the light impinging on an observer. Since this function describes everything that can be seen, I will call it the Plenoptic function (from *plenus*, complete or full and *optic*)." (Adelson and Bergen, 1991; italics are their own)

Photopic vision is a function of reflected light (luminance), therefore the basis for the plenoptic function is the "pencil," which is the mathematical term for the set of light rays passing through any point in space. (Adelson and Bergen, 1991) The authors borrowed an experiment from Leonardo Da Vinci as a paradigm to explain the parameters of the function. They wrote, "Consider, first, a black and white photograph taken by a pinhole camera. It tells us the intensity of light seen from a single viewpoint, at a single time, averaged over wavelengths of the visible spectrum. That is to say, it records the intensity of the distribution P within the pencil of light rays passing through the lens. This distribution may be parameterized by the spherical coordinates, $P(\theta, \phi)$, or by the Cartesian coordinates of a picture plane, $P(x,y)$. A color photograph adds some information about how the intensity

varies with wavelength λ, thus $P(\theta, \phi, \lambda)$. A color movie further extends the information to include the time dimension t: $P(\theta, \phi, \lambda, t)$. A color holographic movie, finally, indicates the observable light intensity at every viewing position, $V_x V_y$ and V_z: $P(\theta, \phi, \lambda, t, V_x V_y, V_z)$. A true holographic movie would allow reconstruction of every possible view, at every moment, from every position, at every wavelength, within the bounds of the space-time-wavelength region under consideration. The plenoptic function is equivalent to this complete holographic representation of the visual world." (Adelson and Bergen, 1991)

As a lead-in to their explanation of the role of early vision in extracting luminous information from the infinite amount available to an observer, Adelson and Bergen offer two propositions:

- Proposition 1. The primary task of early vision is to deliver a small set of useful measurements about each observable location in the plenoptic function.
- Proposition 2. The elemental operations of early vision involve the measurement of any local change among various directions within the plenoptic function. (Adelson and Bergen, 1991)

The small set of useful measurements, detecting local change among various directions describes the mathematical directional derivative or what the authors suggest are "feature detectors."(Adelson and Bergen, 1991) When considered in very small neighborhoods within the seven dimensions of the plenoptic function, the directional derivative might seem too rough a calculation, possibly resulting in a visual world of random noise from uncorrelated measurements. To overcome arguments of this sort, Adelson and

Bergen suggest that the local average derivative of the function is taken, allowing correlated measurements from all dimensions of the function simultaneously.

To apply plenoptic theory to human visual processes, the authors simplify their explanation to a level which conveniently coincides with the static imagery utilized in this thesis. They explain, "At any given moment, a human observer has access to samples along five of the seven axes of the plenoptic function. A range of the x and y axes are captured on the surface of the retina; a range of the λ axis is sampled by the three cone types; a range of the t-axis is captured and processed by temporal filters; and two samples from the V_x axis are taken by the two eyes." (Adelson and Bergen, 1991) Head motion, which would account for the V_y and V_z samples are not considered in their discussion and also not in this thesis, therefore the function simplifies to $P(x, y, \lambda, t, V_x)$.

The authors elaborate on the physiology of human vision and the particular receptor sites at work gathering information in the five axes from the pencil of rays entering the pupil. Although most of this discussion is beyond the scope of this thesis, there are some salient observances made. They note that there are more spatial (x,y) receptor fields in the visual cortex than of any other type and that spatial analysis is the most detailed of all, more occuring in the fovea than on the periphery. From this observance, the authors presume that spatial information is more important to human vision than any other dimension sampled.

The wavelength dimension, λ, is particularly interesting due to its extreme relevance to early color vision and, therefore, this thesis. It is important to note that 'opponent' colors

are from ends of the wavelength spectrum where they do not mix. The authors note that the three human cone types are tuned to only three points on the wavelength axis, one red, one green, and one blue. Figure 19 is used by the authors to present the plenoptic function's reception of the averaging (achromatic), first derivative (blue-yellow opponency) and second derivative (red-green opponency) color information.

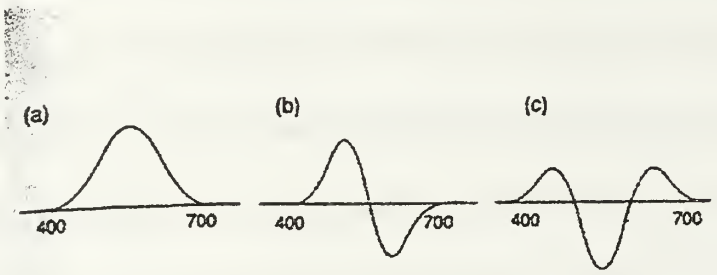

Figure 19. Color information as recieved by the plenoptic function (in nanometers): a) achromatic, b) blue-yellow opponency and c) red-green opponency. (Adelson and Bergen, 1991)

With the five axes of the plenoptic function then, Adelson and Bergen suggest that a "local energy measure" can be assessed without specifying an element (or structure) as mentioned earlier. They wrote, "One may wish to know, for example, that there exists an oriented contour without specifying whether it is an edge, a dark line, or a light line." (Adelson and Bergen, 1991)

Adelson and Bergen reiterate that early vision utilizes the local, low order derivatives, of the plenoptic function to sample a wide range but yet a small sample of the visual information available to the pupil of the human eye. This basic model and the

following one for texture are key to understanding how humans possibly assimilate visual information from the imagery used in this thesis.

4. Visual Texture Segmentation Model

Another visual process that is significant in early scene interpretation and therefore of interest to this thesis is visual texture segmentation. Figure 20 is an illustration used by Bergen and Landy (1991) to introduce the concept of visual texture segmentation. In discussing the figure, they point out the ease at which the rectangular area of X-shaped stimuli is segregated from the L-shaped ones and how the same is not true for the rectangular area of T-shaped stimuli (right). The authors use this simple difference between ASCII stimuli to distinguish "preconscious and rapid" texture segregation from the more deliberate task of pattern discrimination. (Bergen and Landy, 1991)

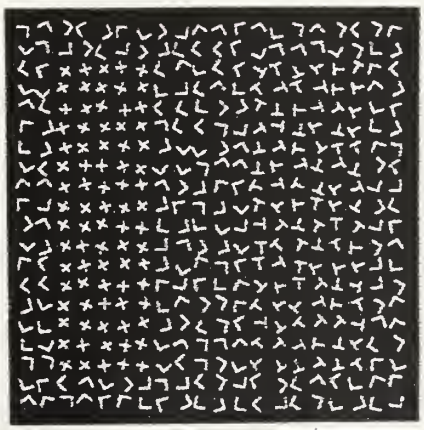

Figure 20. Texture segmentation with ASCII characters. The rectangle of x's (left) pops out while the rectangle of t's does not. (Bergen and Landy, 1991)

In discussing the relationship of texture segmentation to human vision of natural scenes, Bergen and Landy wrote, "Pure texture-based segregation is not a very important phenomenon in everyday visual experience. Objects are not usually distinguished from their backgrounds purely by textural differences. In this respect, the study of pure textural differences (in the absence of differences in brightness, color, depth and other properties) is analogous to the study of isoluminant color differences, which also are not very common in natural scenes. The relative rarity of isoluminant color discrimination in the real world does not imply that color perception is an unimportant component of seeing. Similarly, the rarity of pure texture differences does not reduce the potential importance of texture perception, especially in the visual processing of complex scenes." (Bergen and Landy, 1991)

In stating the motivation for their 3-stage computational model of texture segregation in early human vision they wrote, "Our goal is to investigate the extent to which texture segregation phenomena are consequences of the structure of early visual processes and the representations computed by them." (Bergen and Landy, 1991) The authors' discussion of the physics of the computational model are presented in a depth and detail that is beyond the scope of this thesis, however, a general overview of its structure and prediction capabilities is warranted.

Figure 21 is used by the authors to illustrate a basic outline of the model's interpretation of texture segmentation in early vision.

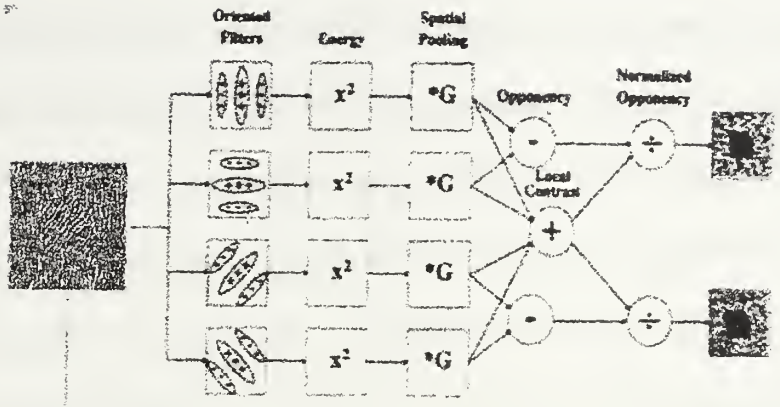

Figure 21. A Basic Model of Early Visual Texture Segmentation. (Bergen & Landy, 1991)

The first column of Figure 21represents a series of input images reduced in spatial resolution by a factor of 2 from level to level (one level shown). The mechanism used for this task was the Gaussian Pyramid algorithm of Burt (cited in Bergen & Landy, 1991), which employs a cascade of linear filters each followed by subsampling to achieve the "blurring" that would otherwise be computationally expensive if done in one step. (Bergen & Landy, 1991) The dendritic for each image in the first column is expanded in the second column to four filters designed to strip-off the respective orientation information from the input they receive. As depicted, the filters sense vertical, horizontal, diagonal left and diagonal right orientation from their input by approximating the second order directional derivative. Since the authors were not interested in the output from the linear orientation filter, they compute in the column labeled "energy" the "local energy" or "the total amount of a particular amount of spatial structure within their region of pooling." (Bergen & Landy, 1991) In discussing the energy calculations they wrote, "We compute energy by squaring

the output of the linear units and then taking a weighted average over a small region. This weighted average is achieved by reducing resolution by a factor of 4 using the same Gaussian pyramid algorithm used to construct the linear filters." (Bergin & Landy, 1991) In the diagram, the Gaussian reduction is done in the column labeled "pooling."

The authors chose to include calculations for orientation "opponency" in the junctions labeled so in Figure 21 (above). They argue that subtracting horizontal from vertical and left diagonal from right, serves computationally to remove any "sensitivity" of the output to the underlying linear orientation filters and to place the output "in quadrature (90° out of phase)" from the two inputs. (Bergin & Landy, 1991) To further separate the opponent signals from any confounding information, Bergen and Landy summed the pooled outputs across all orientations and called this 'local contrast.' They then divided each opponency output by the local contrast, separating the structure information from the contrast information and thereby "normalizing" the output. (Bergin & Landy, 1991) The output, then, is pure luminous information about the stimulus.

With the orientation model described, Bergin and Landy present several examples of texture experiments and the model's performance in detecting orientation-based texture. The one example of particular interest in this thesis is an experiment involving natural textures. Figure 22 is the "straw framed in tree bark" stimulus (Brodatz, cited in Bergin & Landy, 1991) in which separating the texture of the hay from the bark is not done preconsciously due to little coherent difference. Employing the model, however, the normalized opponent outputs allow automatic texture segmentation because the confounding

contrast information has been removed. This result is significant to this thesis because it illustrates the strong impact of local contrast in masking texture information in natural scenes.

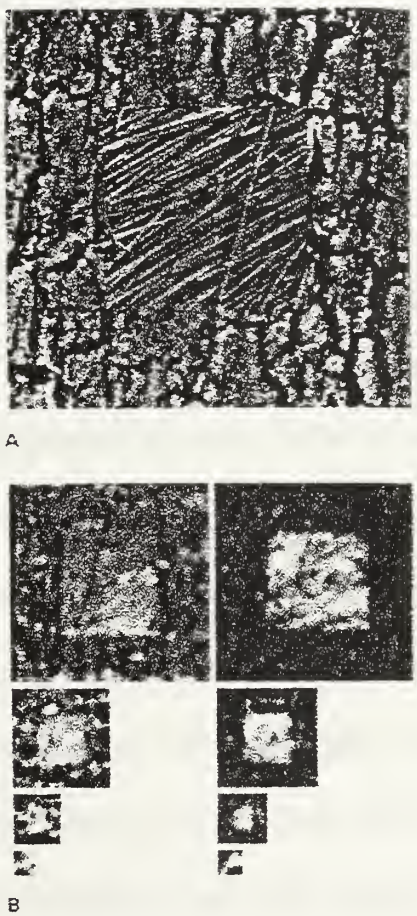

Figure 22. The "straw framed in tree bark" stimulus, A (Brodatz, cited in Bergen & Landy ,1991), and the Bergen-Landy model output, B, conceptualizing how texture segmentation by human vision is accomplished through filtering out confounding contrast information. (Bergen and Landy, 1991)

Two studies in target detection (Wolfe, 1994b, Biederman, et al, 1973), also based on Neisser's work, focused on naturalistic (computer generated) stimuli or natural (photographic) stimuli. Because these studies are fundamental to understanding the methods and conclusions of this thesis, they will be discussed in the following subsections.

5. Guided Search In Naturalistic Stimuli

As described in the previous sections, basic literature in visual search and target detection has evolved to include two stages of cognitive processing by the human visual system. What is not detected preconsciously or in parallel is detected in "a serial, self-terminating search through virtually the entire set of items." (Wolfe, 1994a) A typical visual search paradigm for detecting these processes consists of ASCII characters as stimuli (in the style of Bergin & Landy, 1991) with either a differing color or orientation of the target character as the dependent (fixed) variable. These experiments are, however, a far cry from a representation of real-world imagery. In order to lend some reality to visual search research, Dr. Jeremy Wolfe constructed "Canal World" which is a computer-based experiment that generates 'naturalistic' overhead terrain images with a target embedded in varying amounts of distractors. (Wolfe, 1994b)

Using the canal world experiment, Wolfe (1994a) found that he could determine parallel and serial visual processing by his subjects. However, as he made the scene more continuous, more natural, reaction times rose enough to destroy the usual slope difference of a factor of two between targets processed in parallel and those processed serially.

One significant finding of this study was that real world imagery is difficult to use

in a parallel versus serial search experiments because the amount of distractors in a whole image cannot be appreciably manipulated (Wolfe, 1994a). Other studies, namely the research of Beiderman and others presented in the next section, provided other techniques by which to analyze target detection and accuracy in NVD stimuli.

6. Guided Search In Natural Stimuli

Studies on the effects of overall coherency of a target's setting on accuracy and reaction time in a search task were conducted by Biederman and others (1973). These studies were unique in that they utilized photographs of naturally occurring scenes in their tasks. Their experiments, using what would be considered crude images compared to today's technology, involved flashing 96 slides of scenes that were 'coherent' (spatially intact) or 'jumbled' (not spatially intact). The original photographs were sectioned (cut) vertically into thirds and then horizontally in half, for a total of six sections each. Half of the slides were left coherent and the other half had one section remaining in its original position while the five remaining sections were jumbled randomly or were replaced by a section from a different scene. Section lines were left in the coherent slides as well as the jumbled ones for uniformity and the image on the projection screen subtended a visual angle of 19 degrees.

Subjects were shown a card with a target from one of the sections for five seconds after which one of the slides was presented until a response was given. Reaction time to determine 1) "yes" the target was present, 2) "no" the target was from the scene but was not present (possible-no) and 3) "no" the target was not from the scene and was not present

(impossible-no) was measured. The results showed increased reaction time for jumbled images across all responses, however the increase for the 'possible-no' responses was on the average .75 sec slower than the 'impossible-no' responses. Biederman and others attributed these increases to disruption of the initial holistic characterization of the stimulus as described by Niesser (1967) in his theory of the multistage processing of images. Furthermore they point to a subject's ability to 'make sense' of the jumbled scene and exit faster for 'impossible-no' scenes than for 'making sense' and then searching for the target in 'possible-no' scenes.

In today's vision research terminology, Beiderman possibly would conclude that the 'impossible-no' responses were a result of rapid parallel or preattentive search while the 'yes' responses were from an initial serial or focused search and the 'possible-no' responses were from a secondary, self-terminating serial search. Also, Biederman and others note the number of sections in a jumbled scene may drain visual processing power, which is consistent with the more recent work of Wolfe (1994a) discussed above.

The past two subsections have been reviews of fundamental studies in target detection designed to test human target detection abilities in the daylight photopic world. With all the aditional variables of sensing optical and IR energy and the limitations of the NVDs already discussed, it is understandable that visual tasks become more difficult in the NVD environment. The next section is a discussion of the key variable behind target detection in general, contrast sensitivity. Examples of NVD imagery used in this section will be actual imagery from the four sensors.

7. Contrast Sensitivity

"Human vision with NVDs is a more complex process because, in addition to the normal visual processes, we now add an electro-optical viewing device. Unlike looking through a pair of binoculars, NVGs and FLIRs do not provide direct viewing of the object. Even though vastly superior to night unaided vision, the NVD image is just an artificial TV screen representation of a scene that is not daylight quality." (MAWTS-1, 1995) Effective NVD images can provide the visual system adequate image information to allow good visual performance at two levels: the level of object (contrast) detection and the level of perceptual organization.

On the first level, Campbell and Robson (1994) proposed that the human visual system is able to detect objects because it senses an image by way of simple patterns of parallel light and dark bars called 'gratings.' These bars vary in width, contrast and orientation so there are infinitely many combinations. They hypothesized that the human visual system has sets of neurons called 'channels' that are tuned to different bar widths. Campbell and Robson's 'multichannel model' relates perception of objects in human vision to 'aggregates' of various pairs of gratings whose contrast contributed enough to the image to stimulate 'sensitive' channels (Sekuler and Blake, 1990). When analyzing these aggregates, one must consider the number of pairs of bars imaged on the retina from a certain distance, or 'spatial frequency.' By measuring the contrast threshold necessary to stimulate these channels across spatial frequencies visible to humans, researchers have derived the Contrast Sensitivity Function (CSF), an example of which is shown in Figure 23.

In more familiar terms, combinations high on the CSF curve correspond to high visual acuity (e.g., unobstructed vision) and combinations low on the curve correspond to low visual acuity (e.g., underwater vision without goggles).

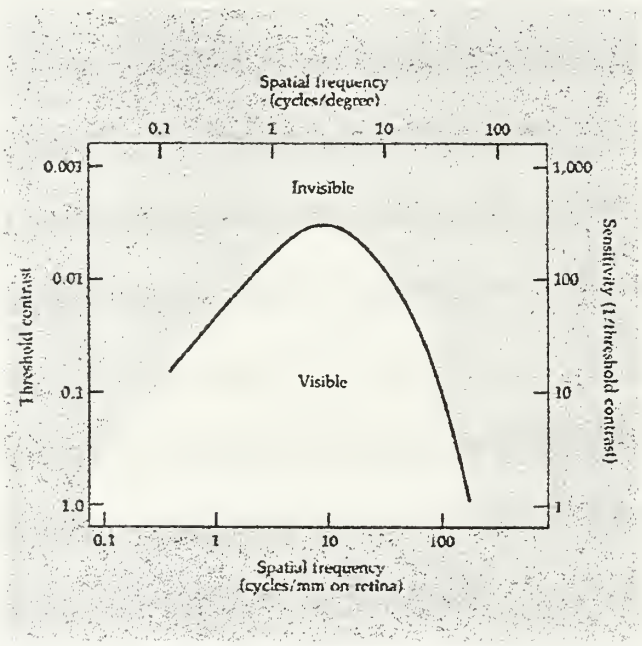

Figure 23. A Contrast Sensitivity Function Of An Adult Human. Visible And Invisible Regions Are Shown According To Spatial Frequency And Contrast. (Campbell and Robson, 1994)

Since NVDs must provide spatial information adequate for good performance on spatial detection tasks, the CSF is an excellent metric for evaluating visual ability while using them. Initial studies at The Center for Visual Science and Advanced Displays (CVSAD), Monterey show that the present NVGs degrade the user's CSF considerably and in a spatially non-uniform manner that is especially detrimental to detection of certain types of image structure (e.g., spatial details and global, low spatial frequency structure). (Krebs,

1994) Additional concerns are a significant reduction of resolution in starlight illumination the effects of blur and a reduction of stereo acuity. The following subsections will be discussions of the contrast information provided by the display of each sensor.

a. I^2 Imagery

Figure 24 is one example of the variable quality of I^2 imagery given a certain set of NVGs and a certain combination of luminance, illumination and atmospheric conditions. This image was taken by the AHPS. Compared to a daytime image of the same scene, the contrast is severely degraded. The degradation is almost enough to elude the contrast sensitivity of an average human, keeping them from detecting the target, a tank truck in

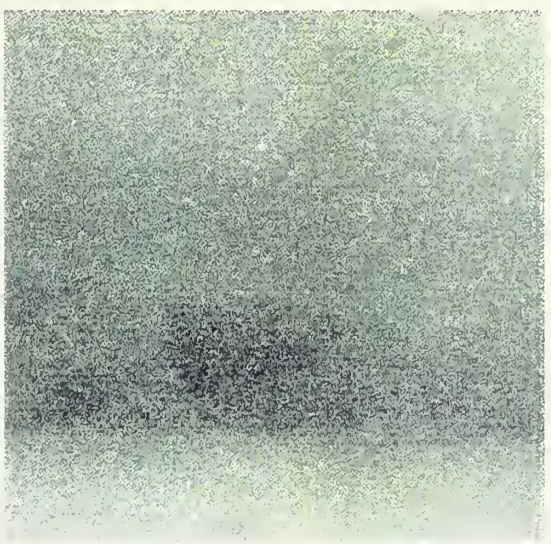

Figure 24. An NVG Image. (Courtesy of NVSED)

the lower right corner. Although some of the target's contrast degradation could be from

shadowing or special paint designed to reduce reflection and therefore albedo, the image as a whole lacks clear borders between field and forest and forest and sky which are important to situational awareness while piloting aircraft at low altitudes (less than 500 feet).

As previously mentioned in the NVD factors section, illumination and luminance are essential to garnering an image from an image intensifier. This image is uniformly poor below the treeline most likely due to less illumination incident to that area. This analysis is made evident by looking at the upper portion of the image and seeing the impact of night sky illumination has on improving the contrast of the image. In the sky, planets, stars and a glow that may be from various phenomena of the night sky are visible. In more illuminated conditions, the objects in this image could be well within the CSF for an average human.

b. IR Imagery

Figure 25 is a FLIR image of the same scene as Figure 24, both were taken simultaneously by the AHPS. The image is a snapshot of the thermal scene that was available given the AHPS' MRTD and a certain combination of emissivities, reflectivities and atmospheric conditions. Compared to a daytime image of the same scene, the contrast is still degraded but in this case more information about the target, background, treeline and sky are available to the user over that of NVGs. It is important to stress, however, that the conditions might have easily been reversed with the NVG image providing more information. Because of the delta T between target and background and the solar heated top of the target with its shaded bottom, the target is more distinct. Also, the warming of the air

by the cooling earth and the relatively low emissivity of the vegetation provide sharp contrast cues about the treeline to enhance navigation and targeting.

Figure 25. An IR Image. (Courtesy of NVSED)

The horizontal temperature bands in the center of the image and above the treeline are examples of areas of homogeneity where an object of the same temperature would not be visible. A good example of this is the absence of a division between foliage of individual trees. Possibly here a deciduous tree or forest with a different emissivity would enable a delta T and a corresponding amount of detectable contrast. Overall, with this particular image, more information is within the human CSF, enhancing situational awareness.

c. Fused Monochrome Imagery

Figure 26 is the result of fusing the images in Figures 24 and 25, processed by the AHPS 'realtime' and available to the pilot virtually instantaneously. This image is an excellent example of the advantages of fusion of NVG and FLIR imagery.

Figure 26. A Fused Monochrome Image. (Courtesy of NVSED).

Although it is not as bright as the FLIR image, the fused image (Figure 26) trades brightness off for more contrast in the foliage, between the field and the foliage and in the night sky. Weighting of the information in each pixel has also kept the target well defined while the natural features are balanced and more textured. Increasing texture lends itself to increased depth perception which is critical to the situational awareness of aviators. Overall, more information is brought into the human CSF with fusion than with the single-band sensors individually. It is important to note here that different NVG and FLIR information input to the fusion algorithm could yeild a completely different image.

d. Fused Color Imagery

Figure 27 is an example of fused color imagery resulting from additional processing of the NVG and FLIR images in Figures 24 and 25 external to the AHPS. This

fusion and coloring technique was performed by the Naval Research Laboratories (NRL) using opponent color contrast (Figures 15-17).

Figure 27. A fused color image. (Image courtesy of NRL)

Most humans have color vision and can appreciate the benefits of contrast made available by color. In Figure 27, the additional contrast is between the field (shades of cyan), the foliage (shades of black) and the night sky (shades of magenta). It is important to stress that the color of a pixel is not necessarily consistent with that under scotopic (daylight) conditions nor will it necessarily be the same from one fused scene to another. Studies by Triesman (1986) and others reveal that color in conjunction with other factors such as shape, reduce reaction time in 'laboratory' target detection experiments. A pilot study, using these and other images, conducted at CVSAD in conjunction with this thesis revealed that experiments with homogeneous NVD images may not reproduce Treisman's

results.

This section was aimed at providing insight into the pros and cons of the imagery output by the four sensors, which is essential to understanding the hypotheses of this thesis presented in the next section.

D. HYPOTHESES

A considerable amount of background information has been presented thus far to explain how the images used in this thesis came about, how humans perceive this information and some possible methods that can be employed to quantitatively assess the impact of the new technology on a visual search task. In light of the hypotheses concerning fusion and coloring, there was an a priori belief that the results from a reaction time and accuracy experiment would favor fusion and color fusion over the IR and I^2 inputs. In order to measure target detection and detection accuracy on imagery from these sensors, the experiment described in the next chapter was designed with the following null hypotheses in mind:

- There will be no difference in mean reaction time across the four sensors. The goal of this hypothesis is to show the alternative is true using analysis of variance on the reaction time results.
- There will be no difference in mean reaction time across the sensor by scene interactions. The goal of this hypothesis is to show the alternative is true using analysis of variance on the reaction time results.
- There will be no difference in mean accuracy across the four sensors. The goal of this hypothesis is to show the alternative is true using analysis of variance on the accuracy results.

- There will be no difference in mean accuracy across the sensor by scene interactions. The goal of this hypothesis is to show the alternative is true using analysis of variance on the accuracy results.

II. METHODS

The experiment construced for this thesis was developed to measure reaction time and accuracy in target detection using real world imagery from the four sensors. Although measuring whether targets in a scene are acquired serially or in parallel from one sensor to another is desireable, manipulating the number of distractors (e.g., adding or subtracting items) in real images and collecting the required volume of images is prohibitive. (Wolfe, 1993) What could be done with natural stimuli, namely moving or removing naturally occurring targets and measuring reaction times in self-terminating searches, was developed (in the style of Biederman, but without jumbling) for this thesis. The methods of this thesis are representative of a recent shift in vision research toward exploring human performance on visual tasks with natural stimuli.

A. EQUIPMENT

The experimental workstation consisted of an 80486 DX2 personal computer equipped with a Texas Instruments TMS340 Video Board and the corresponding TIGA Interface to Vision Research Graphics© (VRG) software. The stimuli were presented on an IDEK MF-8521 High Resolution color monitor (21" X 20" viewable area) equipped with an anti-reflect, non-glare, P-22 short persistance CRT. Pixel size was .26' horizontal by .28' vertical, 800 X 600 square pixel resolution and the frame rate was 98.9 Hz. Brightness of the monitor was linearized by means of and 8-bit look-up table (LUT) for the red, blue and green guns. Responses were recorded on the number pad of a standard (IBM clone)

keyboard. The monitor and keyboard were placed on separate desks with a black cloth draped over both to prevent surface glare. Mesopic viewing conditions were maintained using a small floor lamp (6.8 cd/m^2 luminance) placed on the floor behind the IDEK monitor. A chair and a chin rest (both adjustable) were provided for subject comfort and to help maintain the appropriate distance and viewing angle.

B. STIMULI

IR, I^2 and fused monochrome stimuli available for this thesis originated from 24 bit 'Digital Snapshots' of FLIR and I^2 video taken in-flight by the Fusion Video Interface of the U.S. Army/ Texas Instruments Advanced Helicopter Pilotage System (AHPS). Due to the close proximity of the two sensors in the AHPS pod and timing synchronization of the two video outputs, snapshots from the FLIR and I^2 video FOV are considered 'optically registered' (identical). (U.S. Army/Texas Instruments, 1993) The experimental design required that the stimuli chosen contain at least one target, identifiable in the IR, I^2 and fused monochrome images. From the available snapshots, three scenes were chosen and labeled: 1) "truck," 2) "rectangle" and 3) "tower." The corresponding targets for each scene were: 1) a tanker truck, 2) a rectangular shipping container and 3) a satellite dish.

Construction of the experimental stimuli began with manipulation of the images using Adobe© Photoshop Illustrator. The images were first cropped to a square 460 X 460 pixel size in order to simulate the more likely square or rectangular image display (output device) in an aircraft. The "marquee"(selection) and "zoom"(magnification) capabilities of Adobe© enabled cropping and target movement to be accomplished with a pixel-to-pixel

pixel-to-pixel match within each image and across sensor types for the same scene. The base image was considered the original, unmanipulated image and, for future data analysis, the position of the target was coded as 1 (complete image file encoding procedures available in Appendix A). In each scene, the target was removed using the "lasso"(capturing) technique in Adobe© and the fill for the target void was taken from pixels neighboring the target and chosen to present the most coherent appearance with the least artifacts possible. With the target dubbed out of the scene, the image was coded position 0. Target position 2 and 3 were created by opening two duplicates of the distractor image and pasting the target in two different, spatially correct positions (avoiding "jumbling" used by Biederman et al, 1973).

The resulting pairs of manipulated FLIR and I^2 images were fused and colored by the Naval Research Laboratory's Optical Science Division. Although this was done in the laboratory for this experiment, available technology will eventually allow this to be provided to the pilot in a realtime display. The net result of taking the original images, manipulating the target, fusion and coloring were 48 stimuli: three scenes presented in IR, I^2, fused monochrome and fused color with four positions of the target described above (i.e., 36 images with target and 12 images without). Reprint permission for all stimuli is contained in Appendix B.

After manupulation, all stimuli were subsequently converted to 8-bit, indexed color, IBM compatible image files for interface with the experimental hardware and software. The mean luminance of the images presented varied from 3.0 cd/m^2 (I^2) to 25.0 cd/m^2 (fused

monochrome) for an average mean luminance of 12.5 cd/m^2. Figures 28-31 were constructed to provide a representative sampling of sensors, scenes and positions from the 48 experimental stimuli.

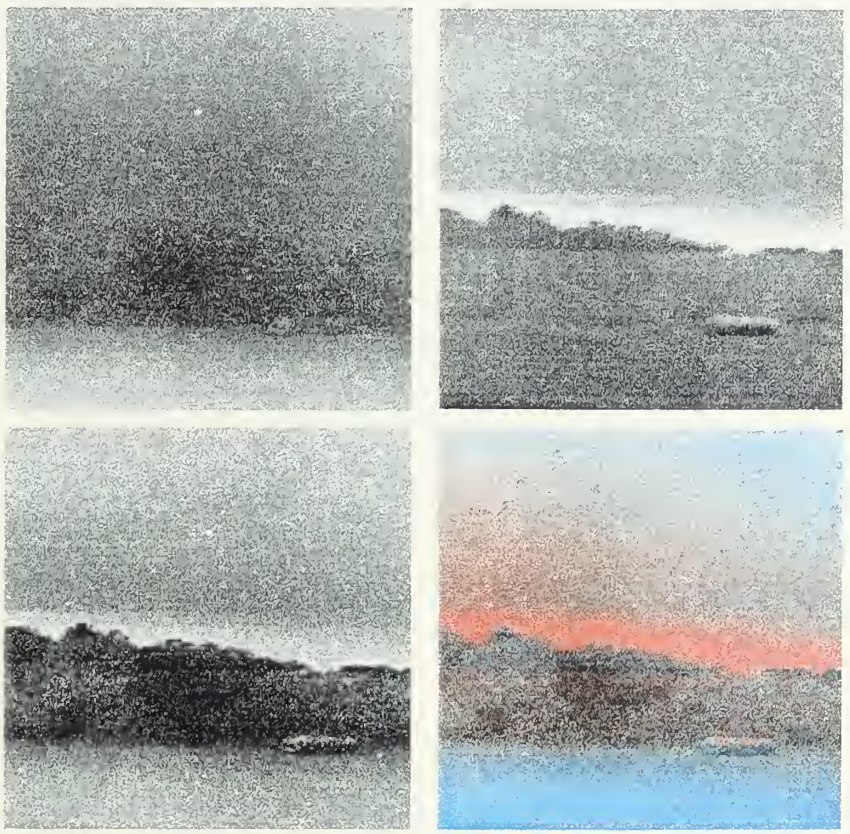

Figure 28. The truck scene as output by the four sensors: I^2 (upper left), IR (upper right), fused monochrome (lower left) and fused color (lower right). (Images courtesy of NRL and NVSED).

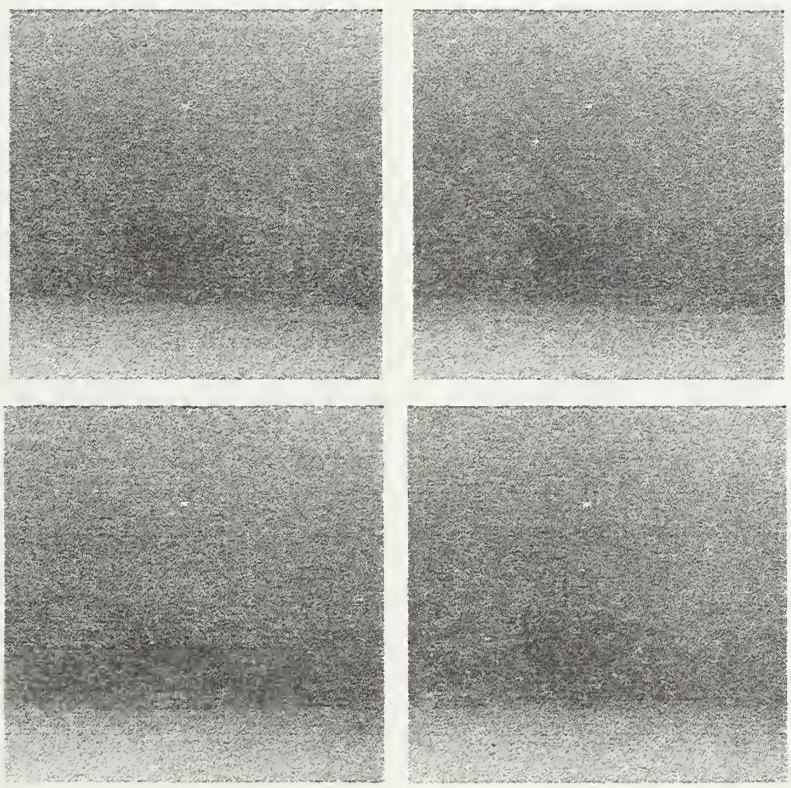

Figure 29. The four positions of the truck scene as presented by an I^2 device: distractor (upper left), position 1 (upper right), position 2 (lower left) and position 3 (lower right). (Images courtesy of NVSED).

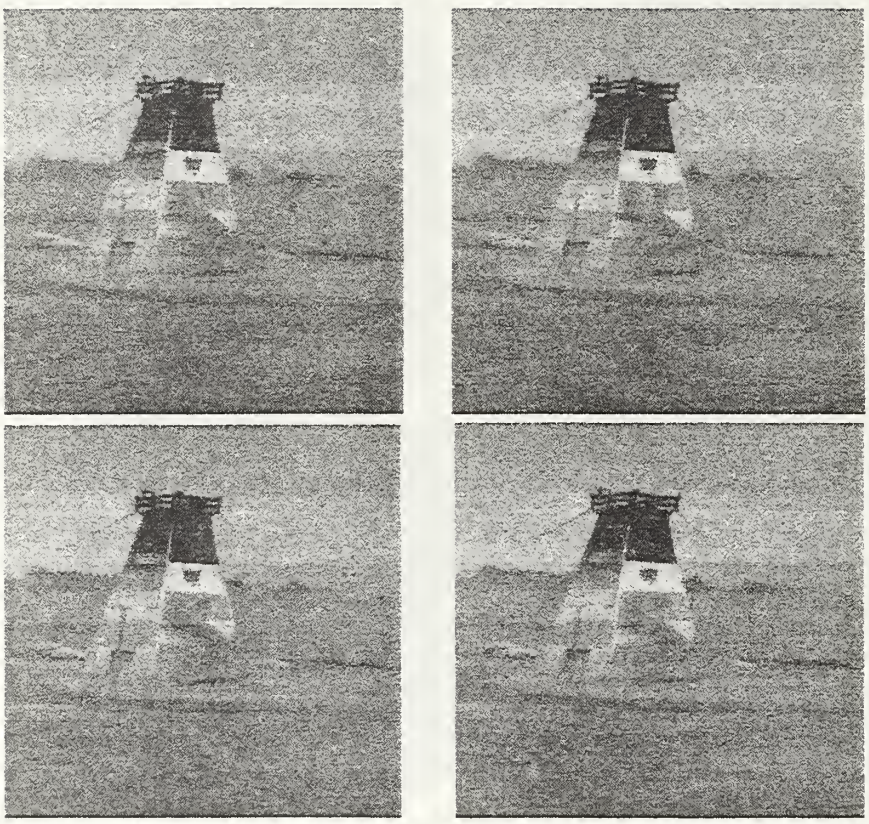

Figure 30. The four positions of the tower scene (satellite dish target) as presented by an IR device: distractor (upper left), position 1 (upper right), position 2 (lower left) and position 3 (lower right). (Images courtesy of NVSED).

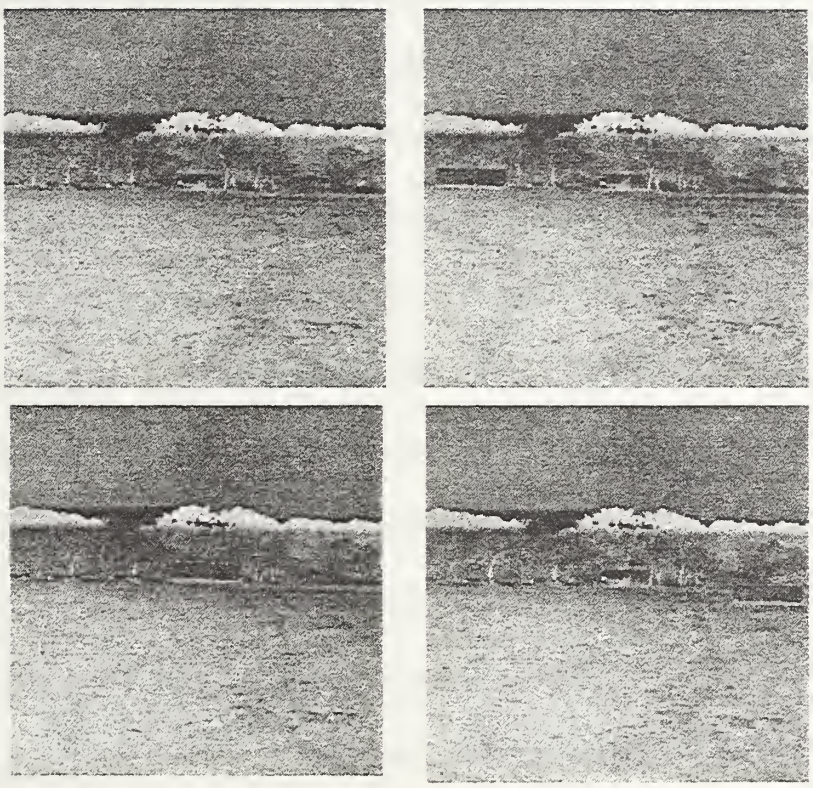

Figure 31. The four positions of the rectangle scene (rectangular box as target) as presented by an fusion device: distractor (upper left), position 1 (upper right), position 2 (lower left) and position 3 (lower right). (Images courtesy of NVSED)

C. EXPERIMENTAL DESIGN

An extension of a randomized block experimental design was employed in the experiment to control nuisance variables without sacrificing the ability to completely explore the stated hypotheses. A randomized block design requires that all subjects receive all treatments randomly. The extension of the design for this experiment involved exposing few subjects to all of the experimental stimuli many times to assist in the blocking and to overcome the vast number of subjects normally required in this type of design. The aim of using this design in this experiment was to 'block' or reduce variability from subject individual differences and, in doing so, focus on the sensor and scene differences (Hayes, 1988). As will be discussed later in the results section, this multiple exposure design may facilitate (as it did here) analysis of the output as a randomized block design as well as a repeated measures design without an appreciable loss of power.

In vision research there are 'targets,' which are the objects of interest, or 'distractors,' which is everything else. For this experiment, images containing the naturally occurring targets described above were considered targets and the images where the target had been extracted were considered distractors. A standard visual search paradigm requires that equal numbers of targets and distractors be presented in an experiment. Accordingly, one matching distractor image for each target image was placed in the theoretical 'urn' of images used for this experiment. In this manner, a total of 36 target stimuli and 36 matching distractor stimuli comprised one 'block' of 72 trials in the experiment, each stimuli drawn randomly and without replacement by the experimental software. A 'session'

of the experiment contained four blocks, the first block was considered practice and the remaining blocks were experimental. Blocks were kept independent by a brief reset procedure between blocks conducted by the software. Each session lasted approximately 30 minutes. The net result of each subject's participation was 648 experimental trials for a total of 3,240 data points. Each subject contributed nine threshold points for the sensor by scene by target/distractor interaction, for a total of 45 threshold points for the experiment.

Stimuli were flashed on the center of the screen in a 10 cm X 10 cm square and were viewed from a distance of 100 cm, therefore subtending a 5.6° x 5.6° visual area on the retina. This visual area is somewhat comparable to what is experienced by users of current Heads Up Displays (HUDs) in military aircraft. An 18 mm X 19 mm white cross-hair, centered on the black screen was employed as a pre-stimulus fixation point. A warning tone (beep) signaled that the stimuli was about to be presented. The stimulus was present until the subject made a selection or until a maximum of 600 ms viewing time had elapsed. The experiment proceeded to the next trial 200 ms after the response was made. A feedback tone signaled an incorrect response for the type of image (target/distractor) that was presented.

D. SUBJECTS

A pilot study aimed at determining if there was a significant improvement in reaction time and accuracy between sensors was conducted at CVSAD. The results of this study were used to determine the number of subjects required to assure at least .80 power under all hypotheses (Appendix C). Using Tang's method it was determined that 5 subjects would be sufficient. Six subjects were chosen to balance the design and to allow for

examination of the a priori assumptions that aeronautic adaptability (having received flight training) and prior NVD use would significantly improve performance. The six subjects used in this experiment were all healthy, male military officers from various services and job specialties undergoing graduate studies at the Naval Postgraduate School, Monterey. Their mean age was 32 years and they all posessed at least 20/20 corrected vision. Half of the subjects were aeronautically adapted (received flight training as part of their job specialty) and, of those three, two had I^2 sensor (NVG) experience. Subjects were naiive to the purpose of the experiment and none had participated in previous visual search experiments. Informed consent was given by each subject. For a more complete listing of subject demographics see Appendix D.

E. PROCEDURE

All subjects completed three sessions, with at least 2 hours between sessions and with no more than 2 sessions completed in a 12 hour period. Before the first session, subjects were read their task instructions and given the opportunity to ask questions. In the instructions, subjects were tasked to rapidly indicate on the keyboard whether they had seen a target in the stimulus (by pressing 1) or no target in the stimulus (by pressing 2). At the beginning of each trial, a fixation crosshair was presented in the center of the screen (Figure 32). The image was presented 200 msec later and the subject commenced their search and made their response. The image was extinguished 600 msec after initial presentation or after the subject made their selection, whichever came first. Reaction time and accuracy scores as well as other pertinent data were collected in text files by the software. Appendix E

provides a detailed description of the collection, collation, enhancement and data analysis methods used.

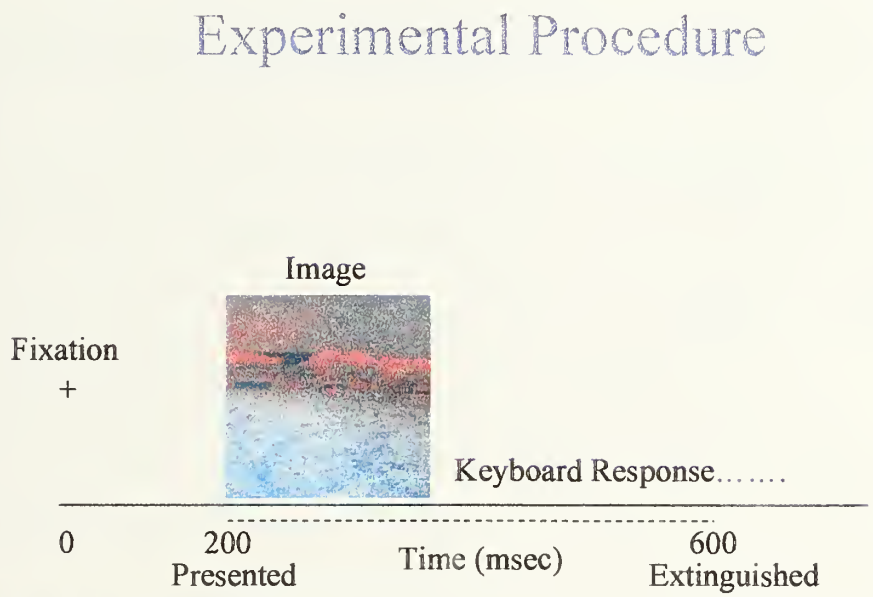

Figure 32. The experimental procedure. A fixation crosshair on the blank screen was followed 200 msec later by the stimulus. The stimulus was extinguished upon subject response or 600 msec, whichever came first.

The VRG program output files enabled gathering each subject's reaction time, accuracy and other parameters for each block of the experiment. With each subject's data collected, collated and placed into a spreadsheet, the analysis was performed using SAS. The results of the analysis are presented in the next chapter.

III. RESULTS

Interviews conducted at the end of each subject's final session revealed that one subject had been daydreaming at times during all three sessions. Visual inspection of that subject's accuracy results revealed an unusually high percentage of errors ($\approx$12%) as compared to the other five subjects and those from the pilot study ($\approx$2%). Inspection of that subject's reaction times showed results as high as 50 seconds (a long daydream) which, based on the pilot study, is unrealistic for these images. Accordingly, this subject's data was discarded from the data set and the analysis was continued.

As previously mentioned, the randomized block (few subjects, many trials) design of the experiment conveniently produced output that could be analyzed using methods for randomized block and repeated measures designs. For both designs, the significance level (α) was set at .05. The results of both designs are presented in the following sections.

A. RANDOMIZED BLOCK ANALYSIS

In the randomized block analysis, reaction time and accuracy data for the nine blocks were collapsed into groupings based on the independent variable(s) selected in the hypotheses. Multivariate analysis of variance (MANOVA) was employed in this design to explore the independent variables and all interactions significant to the dependent measures, reaction time and accuracy simultaneously. The analysis revealed a significant main effect for independent variables sensor (Wilk's Lambda, $F(6,6398) = 13.74$, $p \leq .0001$), scene

(Wilk's Lambda, $F(4, 6398) = 144.53$, $p \leq .0001$) and target (Wilk's Lambda, $F(2, 3199) = 319.95$, $p \leq .0001$). Factorial analysis of the effects between the independent variables revealed there was a significant effect for sensor by scene (Wilk's Lambda, $F(12, 6398) = 23.39$, $p \leq .0001$), scene by target (Wilk's Lambda, $F(4, 6398) = 73.42$, $p \leq .0001$), sensor by target (Wilk's Lambda, $F(6, 6398) = 4.45$, $p \leq .0002$) and sensor by scene by target (Wilk's Lambda, $F(12, 6398) = 2.14$, $p \leq .011$).

With the multivariate analysis complete and the significant interactions noted, the a priori hypotheses and some interactions could be explored using univariate analysis on reaction time and accuracy separately. (Amick & Walberg, 1975) ANOVA on the dependent measure, 'reaction time', showed significant main effect for subject, $F(5, 3200) = 180.63$ $p \leq .0001$, for sensor, $F(3, 3200) = 24.92$ $p \leq .0001$, for scene, $F(2, 3200) = 297.43$ $p \leq .0001$ and for target/distractor, $F(1, 3200) = 612.94$ $p \leq .0001$.

Figure 33 was constructed to assist in exploring the first null hypothesis of this thesis. The mean reaction time (and standard deviation) for the fused color images was 822.06 msec (σ=329.03 msec); for fused monochrome, 787.08 (σ=271.61msec); for infrared, 846.00 (σ=358.31msec); and for I^2, 757.15 (σ=246.15 msec). The ANOVA results and Figure 33 clearly support a significant difference in mean reaction time across the individual sensors, therefore the null hypothesis is rejected.

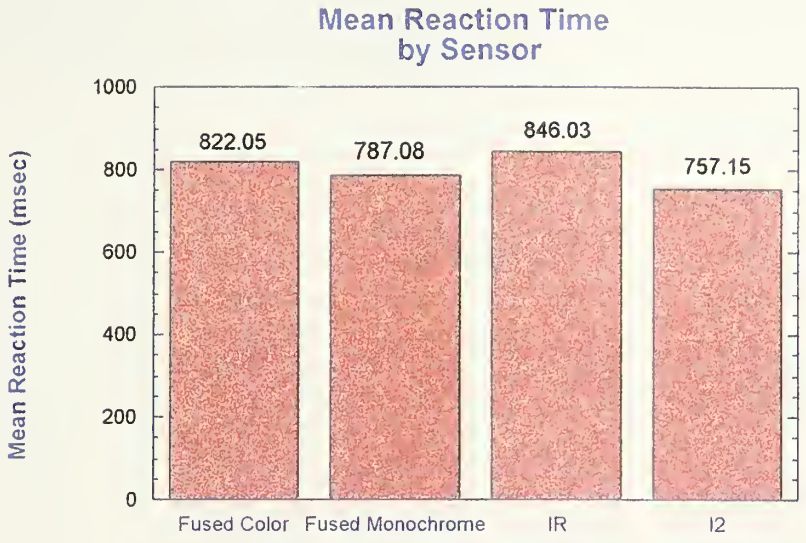

Figure 33. Mean reaction time by sensor ($\underline{F}(3, 3887) = 24.92$ $\underline{p} \leq .0001$). I^2 images yielded the lowest mean reaction time while IR yielded the highest. Of the fused images, fused monochrome yielded the lowest mean time.

ANOVA on the dependent measure, 'accuracy,' showed significant main effect for subject, $\underline{F}(5, 3200) = 38.46$ $\underline{p} \leq .0001$, for scene, $\underline{F}(2, 3200) = 6.81$ $\underline{p} \leq .0011$ and for target $\underline{F}(3, 3200) = 12.79$ $\underline{p} \leq .0004$.

Figure 34 was constructed to assist in exploring the second null hypothesis of this thesis. The mean accuracy (and standard deviation) for fused color images was 99.4 percent (σ=0.078 percent); for fused monochrome, 98.3 percent (σ=0.125 percent); for infrared, 97.7 percent (σ=0.147 percent); and for I^2, 98.1 percent (σ=0.134 percent). The ANOVA results and Figure 34 do not support a significant difference in mean accuracy across

sensors, therefore this null hypothesis cannot be rejected. In these images, there was no significant mean accuracy difference across sensors.

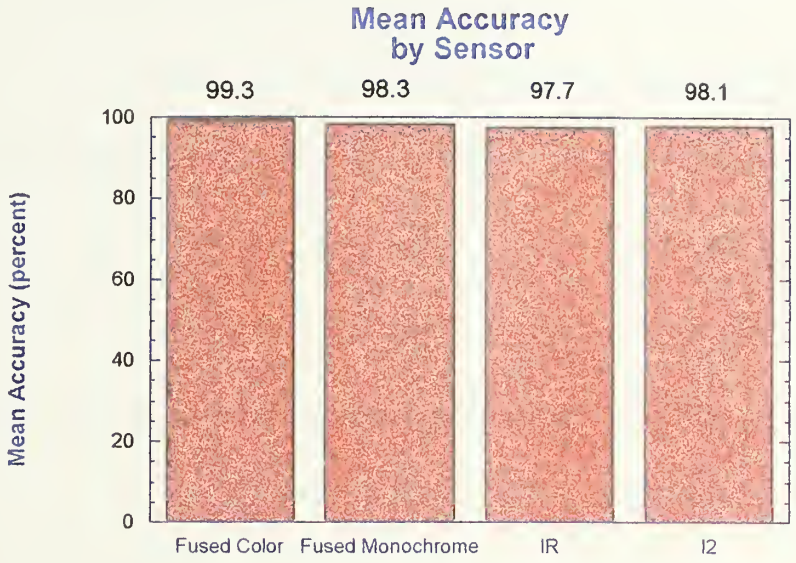

Figure 34. Mean accuracy by sensor ($F(3, 3200) = 2.53$ $p \leq .0554$). Fused color images yielded the highest mean accuracy while IR images yielded the lowest. The relatively small difference in accuracy across sensors made this measure insignificant.

Figure 33 illustrates that the lowest mean reaction time came from the I^2 images with fused monochrome, fused color and IR following in order . Figure 34 illustrates the fact that the accuracy results do not mirror the reaction time results, fused color having the highest accuracy and fused monochrome, I^2 and IR having essentially the same percentage of errors. Tukey Groupings for dependent measure, 'reaction time,' showed all sensors were significantly different except fused color and IR. Tukey Groupings for dependent measure,

'accuracy,' showed that those same two sensors (fused color and IR) were the only ones significantly different. These results were surprising at first, because the full impact of the treatments on these images and on the visual search process was not understood. More in-depth analysis of possible interactions between treatments was needed.

ANOVA on dependent measures reaction time and accuracy for the following factorial interactions yielded the corresponding results: sensor by scene (reaction time: $\underline{F}(6, 3200) = 44.05$ $\underline{p} \leq .0001$), accuracy: $\underline{F}(6, 3200) = 5.23$ $\underline{p} \leq .0001$), scene by target/distractor (reaction time: $\underline{F}(6, 3200) = 142.65$ $\underline{p} \leq .0001$), accuracy: $\underline{F}(6, 3200) = 4.93$ $\underline{p} \leq .0073$) and sensor by scene by target/distractor (reaction time: $\underline{F}(6, 3200) = 2.15$ $\underline{p} \leq .0447$), accuracy: $\underline{F}(6, 3200) = 2.2$ $\underline{p} \leq .0399$). Figures 35 through 40 were constructed to assist analyzing these interactions.

Figure 35 illustrates the sensor by scene interaction effects for reaction time, which is the basis for the third null hypothesis. The mean reaction times are roughly parallel across sensors for the truck and the rectangle scenes but they are highly variable in the tower scene. Visual inspection of the tower images revealed that the target is harder to find when the image is from IR or fused color sensors, otherwise the mean reaction times for the I^2 and fused monochrome images are almost equal to those of the corresponding rectangle images. The ANOVA results and Figure 35 clearly support a significant difference in mean reaction time for sensor by scene, therefore the null hypothesis is rejected.

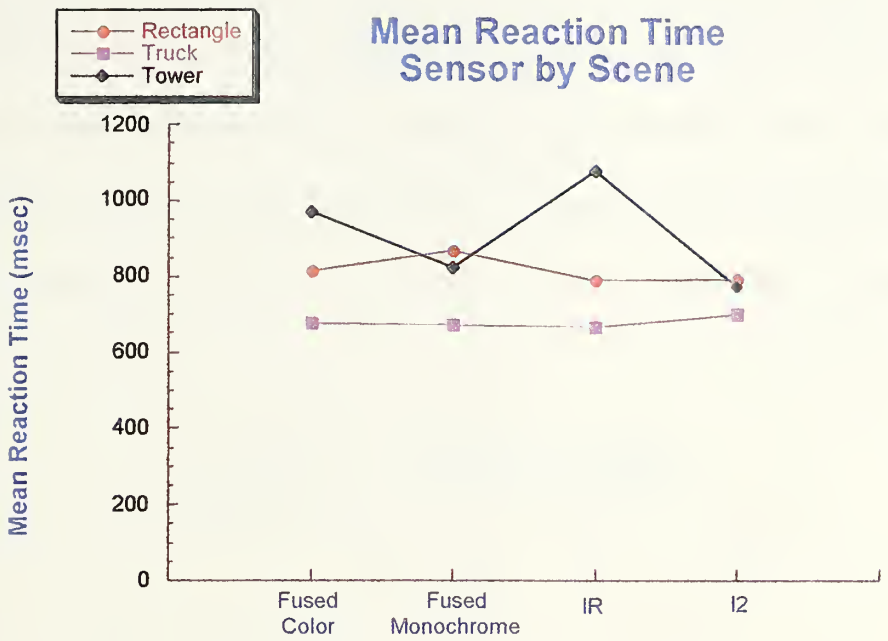

Figure 35. Mean reaction time, sensor by scene ($F(6, 3200) = 44.05$ $p \leq .0001$). The rectangle and truck scenes are roughly parallel across sensor, with a 100 msec split between each scene. The tower image displays high variability with fused color and IR scenes roughly 200 msec higher than fused monochrome and I^2.

Figure 36 illustrates the sensor by scene interaction effects for accuracy, which is the basis for the fourth null hypothesis. A one percent decrease in accuracy from fused color to I^2 for the truck scene is representative of the decreasing amount of global information across the sensors for this scene. The corresponding reaction times for the truck scene in Figure 35 illustrate that the decrease in global information was not significant enough to drive reaction time up across the same sensors. Visual inspection of the truck images reveals that the global and local information available is good across all sensors. The high variability in accuracy for the rectangle and tower images (three and four percent

respectively), almost mirrors the variability in reaction time for the same sensors (Figure 36). The longer search times and sometimes higher error rates are consistent with the decrease in global and local information in the rectangle and tower scenes, which is consistent with the literature. The ANOVA results and Figure 36 clearly support a significant difference in mean accuracy for sensor by scene, therefore the fourth null hypothesis is rejected.

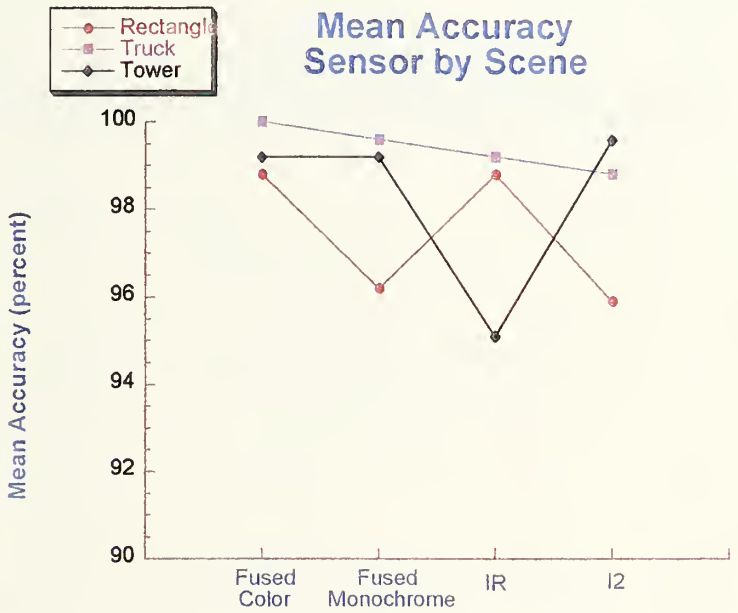

Figure 36. Mean Accuracy, sensor by scene ($F(6, 3200) = 5.23$ $p \leq .0001$). The truck scene shows a one percent decrease across sensors, the rectangle scene decreased roughly three percent for the fused monochrome and I^2 sensors and the tower scene decreased roughly four percent for the IR sensor.

Post hoc analysis on factorial interactions beyond the a priori hypotheses was conducted to explore other possible effects on the data. For instance, Figure 37 illustrates the scene by target/distractor effects for reaction time. According to visual search literature,

a search in the distractor scene for the target should take longer, ending when the subject is satisfied the target is not present (self-terminating). In Figure 37, the truck and rectangle images display roughly the same 100 msec extra searchtime required for subjects to self-terminate their search. In the tower image, with more clutter (natural distractors), subjects required, on the average, 400 msec extra search time in the distractor.

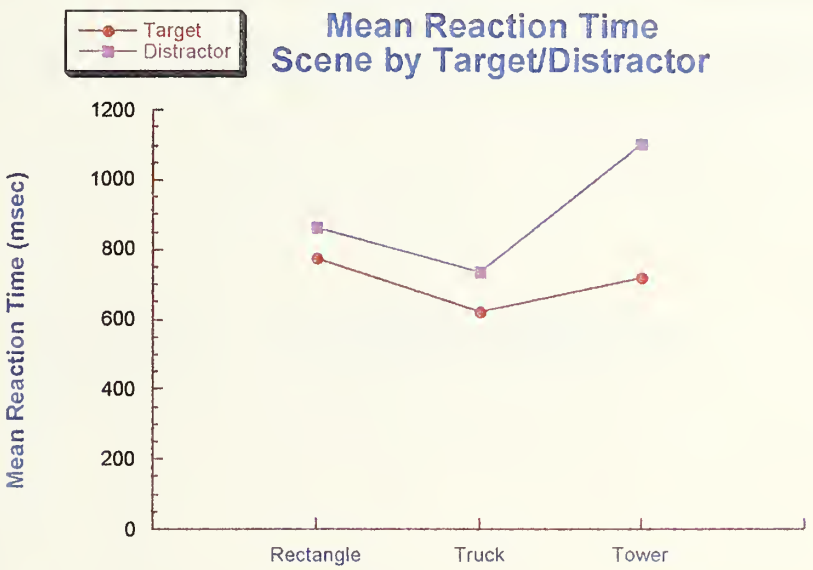

Figure 37. Mean reaction time, scene by target/distractor ($F(6, 3200) = 142.65$ $p \leq .0001$). The truck image had the lowest pair of reaction times with a 100 msec split between target and distractor. The rectangle scene also had a 100 msec split but at a higher reaction time. The tower scene had the widest split, 400 msec, between target and distractor.

Figure 38 illustrates the scene by target/distractor effects for accuracy. According to visual search literature, distractor points should plot slightly above the target points for

the same scene (reflecting higher accuracy with a self-terminating search). In Figure 38, however, the truck scene departs from convention as its target images display slightly higher accuracy than its distractor images. This departure, matched with the relatively low reaction times for target and distractor by scene (Figure 37) and the results for the truck image in Figures 37 and 38, is a strong analytical indication that the truck image was possibly 'too easy' for the task. The results suggest that, on the average, it was slightly easier for the subject to correctly identify the presence of the truck target with less reaction time than required to correctly identify its absence. According to the literature, subjects involved in

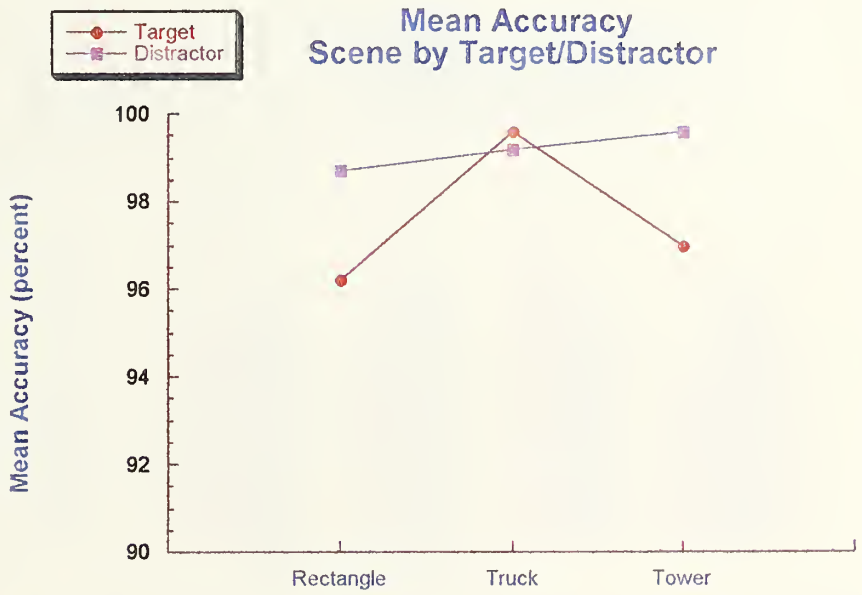

Figure 38. Mean accuracy, scene by target/distractor ($F(6, 3200) = 4.93$ $p \leq .0073$). The rectangle and tower scenes exhibit roughly a two percent split in reaction time while the truck scene exhibits almost equivalent accuracy for targets and distractors.

a self-terminating search, would normally have a higher error rate when the target was present in a scene.

A summary of the significant factorial interactions is provided in the sensor by scene by target/distractor graphs in Figures 39 and 40. By focusing on the pairs of bars with equivalent markings, one can visualize all the interactions with regard to reaction time (Figure 39) and accuracy (Figure 40). For example, in Figure 39, all the right-hand (distractor) bars in the pairs are taller than their left hand (target) counterparts, signifying longer reaction times for a self-terminating search. Also, the spread between target and distractor bars is always largest for the tower scene, signifying the presence of more

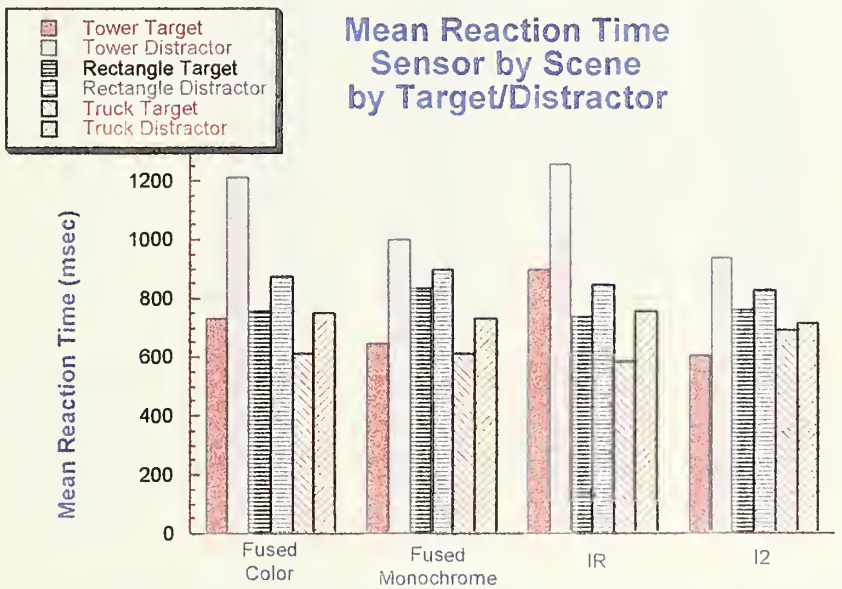

Figure 39. Mean Reaction Time, Sensor by Scene by Target/Distractor ($F(6, 3200) = 2.15$ $p \leq .0447$). Factorial Interactions Are Visualized By Comparing Pairs of Bars Within a Sensor Group and Between Sensor Groups.

information than the other scenes. While the spread between reaction times of targets and distractors for the truck and the rectangle scene are roughly equivalent, the pairs for the truck image are always the lowest, signifying the simplicity or lack of clutter in the scene.

Figure 41 illustrates a summary of the factorial interactions with regard to accuracy. Visible in this graph is the overall high accuracy percentage except for the IR tower target scene and the I^2 rectangle target scene, signifying relatively 'harder to find' targets for those

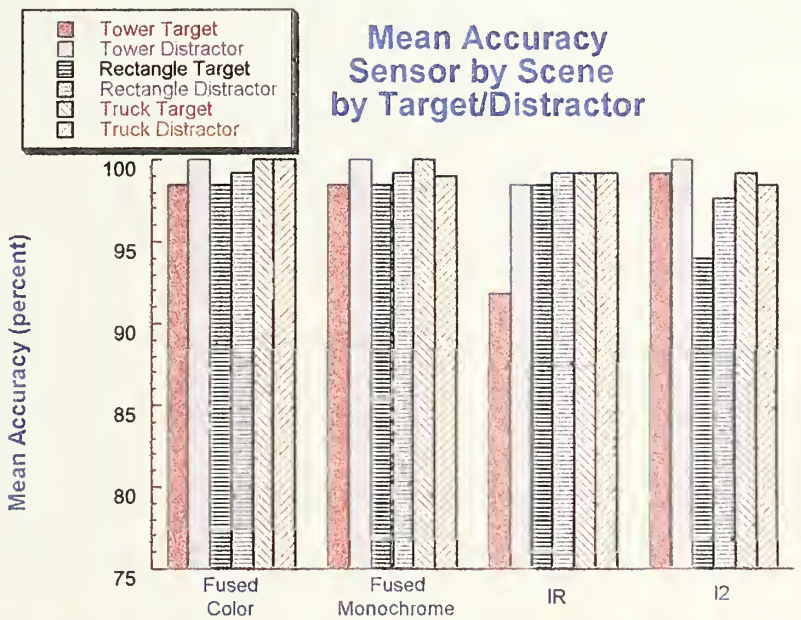

Figure 40. Mean Accuracy, sensor by scene by target/distractor ($F(6, 3200) = 2.2$ $p \leq .0399$). Factorial Interactions Are Visualized By Comparing Pairs of Bars Withing A Sensor Group and Between Sensor Groups.

scenes. Also visible is the inversion of the target accuracy over the distractor accuracy in the IR and I^2 truck target scenes, highlighting a scene that is possibly too simple for this experiment.

B. REPEATED MEASURES ANALYSIS

Post hoc analysis on any possible effects arising from this experimental design (few subjects, many trials), from learning (the first block as training) or from fatigue were explored using a repeated measures analysis. In the repeated measures analysis, with 'block' as the repeated measure, reaction time and accuracy data for the nine blocks were not collapsed as they had been for the randomized block design. In this analysis, as with all repeated measures designs, multivariate analysis of variance (MANOVA) was employed to explore the independent variables and interactions significant to the dependent measures, reaction time and accuracy, as the dependent measure, block, progressed from one to nine. With learning, one expects an increase in accuracy across blocks with a corresponding decrease in reaction time, therefore MANOVA could not be performed on reaction time and accuracy simultaneously as in the randomized block analysis.

'Within-subject' ('within-block' here) analysis on the dependent measure reaction time revealed that there was a significant main effect (Wilk's Lambda, $F(8, 266) = 58.96$, $p \leq .0001$). Within-subject analysis on the dependent measure accuracy revealed that it also was significant (Wilk's Lambda, $F(8, 266) = 5.91$, $p \leq .0001$). Between-subjects analysis on dependent measure reaction time revealed a significant main effect across independent measures sensor (Wilk's Lambda, $F(24, 772) = 1.74$, $p \leq .015$), scene (Wilk's Lambda, $F(16, 532) = 1.72$, $p \leq .0385$) and target/distractor (Wilk's Lambda, $F(40, 1162) = 2.16$, $p \leq .0001$). Between subjects analysis on dependent measure accuracy revealed no significant effects across the independent variables. Figures 42 through 51 were constructed to assist

in visualizing any repeated measures trends.

Figure 42 and Figure 43 illustrate the within subjects effects for block on reaction time and accuracy respectively. The trends visible are representative of learning with time

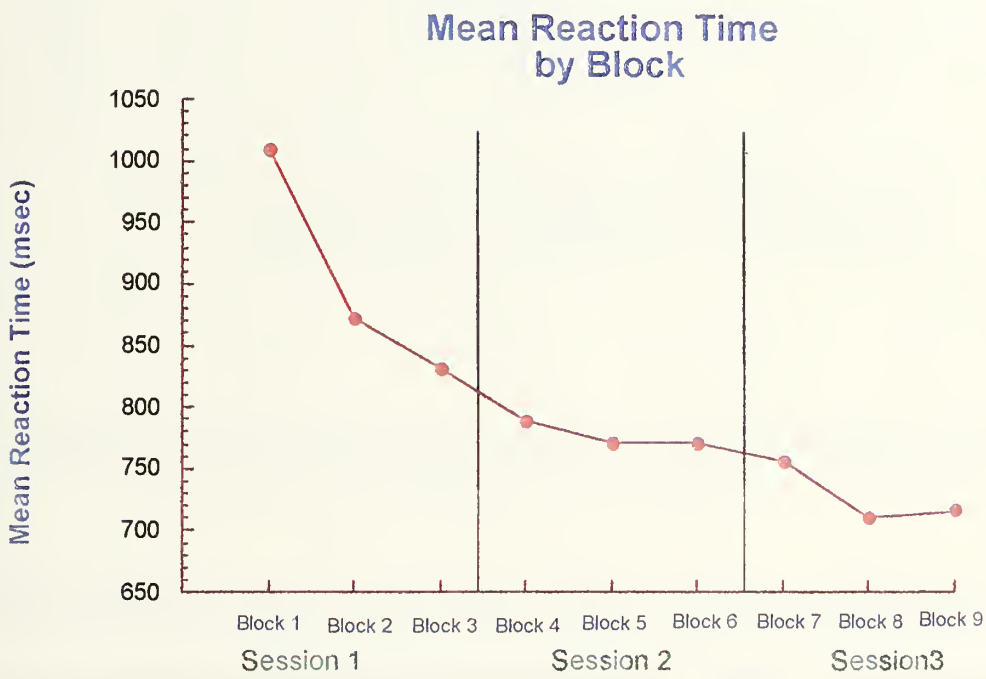

Figure 41. Mean reaction time by block. Reaction time decreases across block except for a 10 msec increase between block 8 and 9.

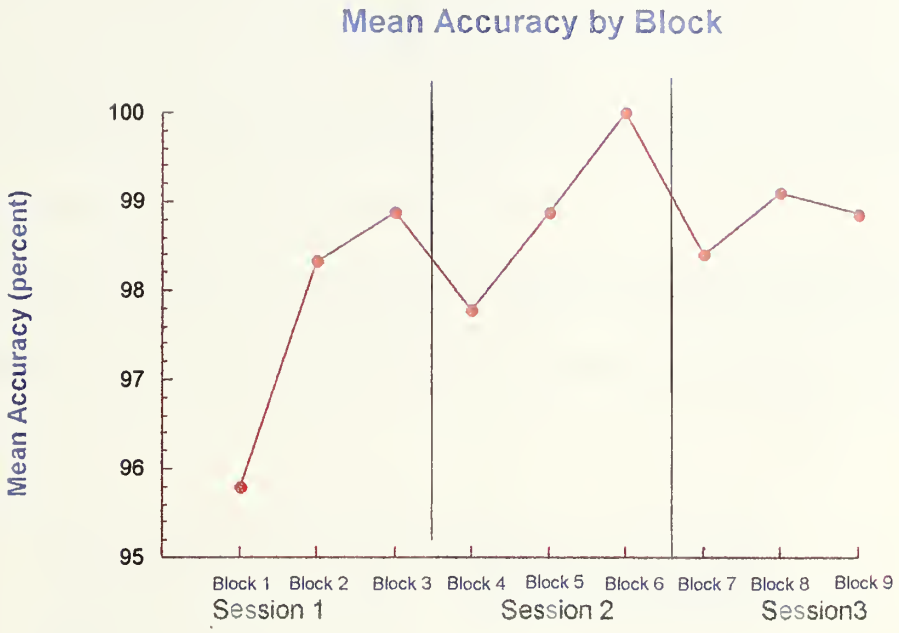

Figure 42. Mean accuracy by block. Accuracy increases across block, except for a 1/4 percent decrease between block 8 and 9, possibly due to fatigue.

as subjects repeat three blocks in each session. What is interesting in these graphs is there appears to be steady improvement (lower reaction time, higher accuracy) as the blocks progress, even though subjects are 'trained' prior to data collection. The departure from the trend from block 8 to 9 is possibly representative of fatigue or complacency.

Figure 44 illustrates the block by sensor trends for dependent variable reaction time. Although all sensors exhibit a downward trend in the first session, the second and third sessions contain blocks where reaction time almost levels off (fused color, block 5) or

spikes upward (IR and I^2, block 5; fused monochrome, block 6; fused color and monochrome, block 9). Since a greater proportion of the increases (4 of 5) occur in the last block of sessions 2 and 3, they are attributed to fatigue. Also visible in Figure 44 is the fact that the I^2 sensor has, on the average, the lowest mean reaction time, which again does not support the fusion and coloring hypotheses.

Figure 45 illustrates the block by scene trends for dependent variable reaction time. Inspection of the graph reveals a large increase in reaction time for the tower scene in block

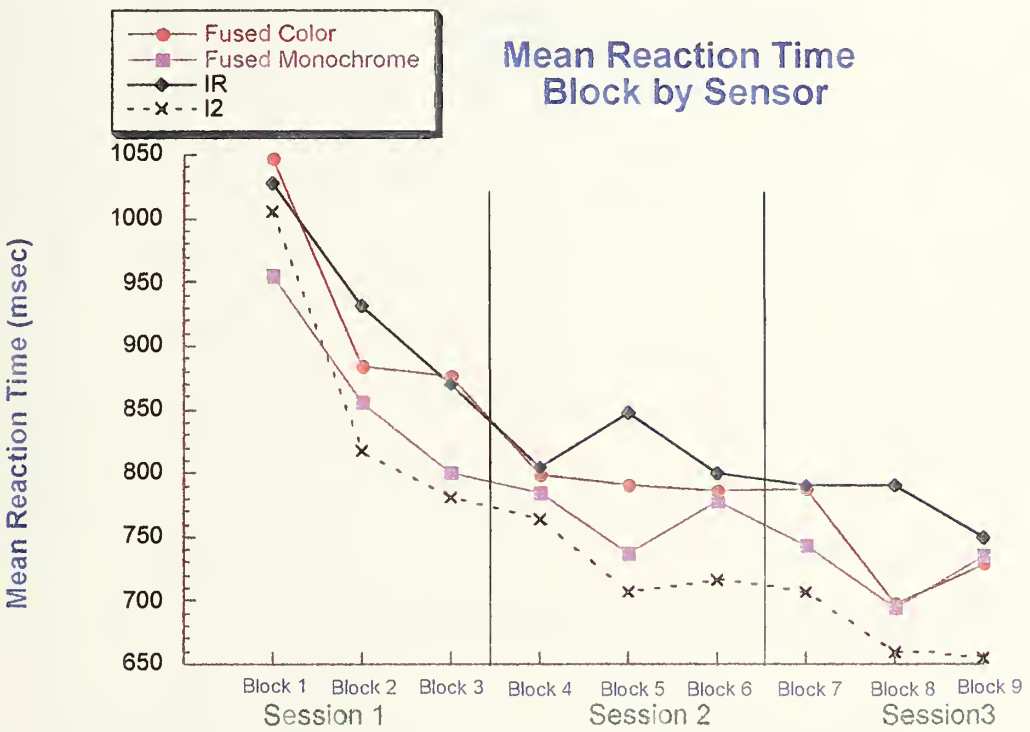

Figure 43. Mean reaction time, block by sensor. All blocks of the first session exhibit a downward trend. In the second and third sessions, I^2, fused monochrome and fused color all exhibit increases in the third block possibly attributed to fatigue.

9 and a slight increase for the truck scene in block 6. Otherwise, all scenes exhibit roughly 200 milliseconds decrease in the first session, almost no change in the second session and mixed changes in the third session. The significant change in the tower scene during block 9 can possibly be attributed to both its complexity as an image and subject fatigue. Also significant in Figure 44 is the large (300 msec) gap between the tower and the truck scene

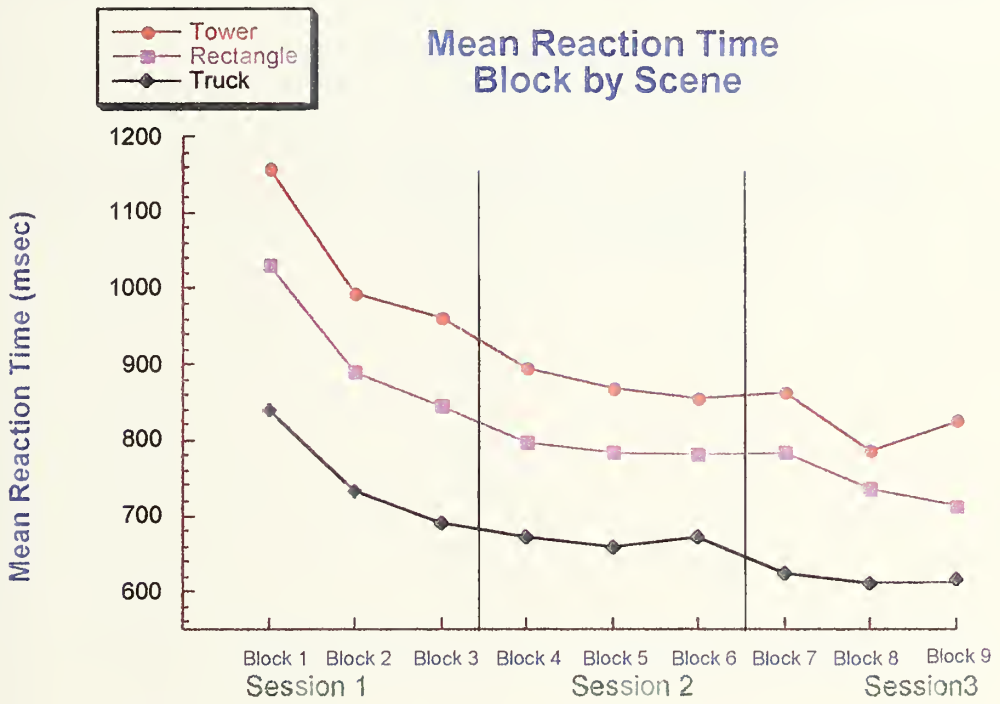

Figure 44. Mean reaction time, block by scene. The tower scene has the highest mean reaction time while the truck scene has the lowest. The tower scene's increase in block 9 is possibly attributable to fatigue.

across blocks with the rectangle in between (closer to the tower scene though). This trend can be thought of as an indicator of difficulty for the experimental scenes: tower, most complex; truck, least complex and rectangle, in between.

For ease of analysis, the block by position interaction for reaction time has been divided into two graphs, Figures 45 and 46. Figure 45 illustrates the steadily decreasing trend in reaction time for the distractor across blocks. This trend is what would be expected as subjects learn and reduce their time to conduct a self-terminating search of the scene. The leveling slope in block 9 is possibly attributable to fatigue and complacency.

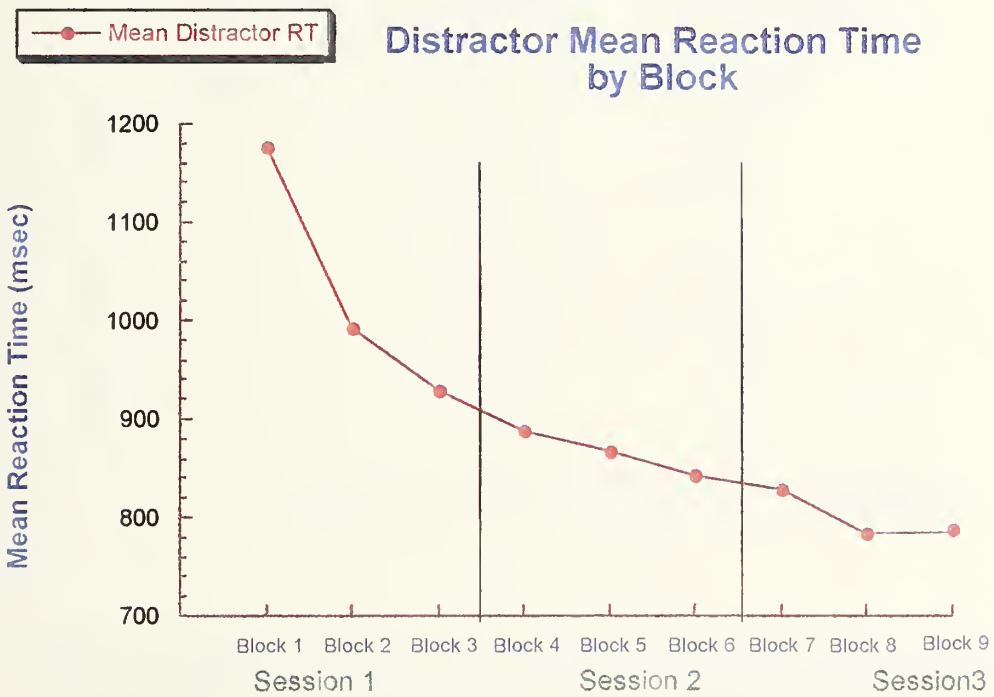

Figure 45. Mean reaction time, block by distractor. A sharply decreasing trend ends level by block 9, possibly due to fatigue.

Since targets were positioned roughly centered in the scene (without losing coherence) or to the left or right of center, Figure 46 has been split into two lines corresponding to target position. Due to their proximity to the location of the prefocus fixation point, the centered targets are expected to yield lower reaction times. Inspection of the figure reveals that the center always is lowest, even when the subjects are tired (block 6 and 9) and performance on the off-center targets has leveled off.

Figure 46. Mean reaction time, block by target position. Centered targets always yield a lower reaction time due to their proximity to the fixation point.

The repeated measures and randomized block analysis above has presented an in-depth look at the factors significant to the experimental stimuli. The next chapter will be a presentation of the conclusions of this research and discussion of possible parallels to vision research literature.

IV. CONCLUSIONS

This thesis was born from research aimed at improving night vision devices by employing the reemerging technology of sensor fusion displays and the new technology of color fusion displays. The experiment designed for this thesis was the first of its type in published visual search literature dealing exclusively with 'natural,' 'coherent' imagery from I^2, IR, fused and fused color displays.

The four hypotheses stated in the introduction were formulated a priori with the belief that the four sensors and the 'raw' (unmanipulated) NVD scenes they generated were unique and warranted exploration as independent variables. There was also an a priori belief that imagery from fusion and coloring would provide superior results in visual search tasks.

The modified experimental design was employed to completely explore dependent measures reaction time and accuracy in target detection, factors which are critical to safe accomplishment of aviation missions. Although there were assumptions about the variability of the data involved in the modified randomized block design, both the randomized block and repeated measures designs provide the same outcome in their ANOVA results - only the structure of the outputs differs.

The robust results and discussion presented in the previous chapter support rejecting all but the second null hypothesis - there was a failure to reject that the mean accuracy across sensors are equal. This failure to reject, the fact that neither the fused nor colored images yielded the lowest reaction time and the fact that the truck scene yielded a

significantly lower reaction time than the other scenes, prompted investigation into possible quantitative and qualitative explanations for these occurrances.

A review of the experimental procedure revealed that the exposure time to the stimuli (600 ms) was longer than what is required to truly test target detection accuracy (hence there was no significant difference between the high mean accuracy scores). Examining the two remaining occurances led to a qualatative comparison of the information content in each image.

While tasks in most standard visual search paradigms are 'too artificial' for use with NVD imagery, the results and the ensuing qualitative comparison of the experimental stimuli did shed some light on relationships between visual search in NVD imagery to the "body of sophisticated theory" that exists regarding laboratory visual search. Two significant contributions to vision research, resulting from the comparison, are noted:

- Fusion and coloring of NVD images greatly impacts global (scene) and local (target) contrast provided by the single-band inputs. In return, the impact on local contrast affects performance on serial, self-terminating tasks.
- Scene content in NVD images (the presence of numerous man-made or natural objects other than the stated target) greatly impacts performance on serial, self-terminating tasks.

These contributions are related to established visual search theories in the discussion below.

The "straw framed in tree bark" modeling (Figure 21) referenced by Bergin and Landy (1991) highlights the possible impact contrast information has in confounding texture segmentation. NVDs provide the viewer contrast information limited by performance of

the sensor and output device. This, combined with effects of fusion (boosting low-pass elements) and coloring (assigning colors according to luminance values) were suspect in decreasing texture segmentation on the target boundaries for the experimental stimuli, and subsequently causing reaction time differences. Close visual inspection of the experimental images from this thesis confirmed this belief.

In the Truck and Rectangle fused monochrome images, visual inspection and data analysis supports excellent IR contrast inputs and poor I^2 contrast inputs producing good global contrast with degraded local target contrast – a tradeoff resulting in increased reaction time from IR to fused monochrome. Truck and Rectangle fused color images also display these characteristics from the inputs and again result in increased reaction time from IR to fused color. A reversal is encountered with the tower scene where scant contrast information for the target in the IR image is combined with good I^2input. Because the exact fusion algorithm used to create the fused monochrome images is not known, one can only speculate that due to the local luminance mean calculation, the fused monochrome image exhibits good global contrast but local target contrast is degraded enough to slow reaction time from I^2 to fused monochrome. In the fused color tower scene, there are good global attributes from the color but the local target attributes are confounded by a lack of color contrast from the background and, therefore, the satellite dish is almost imperceptable.

As stated in the introduction, 'natural' or 'real-world' stimuli do not easily lend themselves to standard visual search experiments which require manipulation of the target and distractors to measure whether preattentive (parallel) or postattentive visual processes

are at work. As Triesman (1985), Adelson & Bergen (1991), Wolfe (1994a) and others have found, the closer a scene gets to being real-world, the less results of standard search paradigms apply. The results of the analysis on mean reaction time and mean accuracy exposed the possibility that the truck scene was possibly not 'hard' enough a scene to use in the experiment. One might compare the task of finding the truck target to that of a lone ASCII character on a blank field. However, the usefullness of this image is evident in analyzing the increase in reaction time across scenes beginning with the truck image and increasing to the tower image.

Inspection of the experimental scenes revealed that there is progressively more information in these images, causing a natural increase in distractors and subsequently reaction time. In this way, the experimental stimuli varied in information content from simple (the truck scene) to more complex (the rectangle scene) to most complex (the tower scene). In a standard visual search paradigm, the increase in complexity would be controlled with more or less ASCII characters or other distractors in the experimental field.

In closing, it is important to note research by others utilizing this type of imagery and to discuss how the contributions of this thesis and their correlation with established visual search theories opens the possibility for additional research. Two studies employing this 'type' of imagery have been conducted concurrently with this experiment. The first study, conducted at CVSAD (Krebs, et al, 1996), was a pairwise comparison task involving 25 images from 5 sensors (I^2, IR, fused and 2 color algorithms) with each image in a pair presented for 3 seconds, the pair separated by a 100 ms interstimulus interval. Fifteen

subjects were tasked to determine which image in the pair best presented the target of interest. Of the five image types, the two color algorithms were selected 'best,' followed by IR and fused monochrome tied for second and I^2 third. Understandably, in this type of aesthetic comparison, human association of 'color' with 'quality' would cause color to be chosen both when it was truly better and also when the comparison was close. Again, this task differed from the methods of this thesis but the results are equally important.

The second study, conducted at the University of Louisville, KY (Essock, et al, 1995), was a pure accuracy task involving 1.5° patches cut out of IR, I^2 and fused color images (the authors note that sensor performance and therefore image quality was lacking). Each session started with training on the target set in the complete, original images. A centered fixation cross was presented for 250 ms, followed by a randomly selected target or distractor patch flashed for 200 ms, followed 20 ms later by a checkerboard mask to terminate visual processing. Ten subjects were tasked to rapidly indicate whether the patch they viewed was a target or not. The results of this study showed the fused color imagery was superior in accuracy with the IR second and I^2 third. The results were significant in determining which imagery provides the best early perceptual organization, however, the quality of the single-band images being poor may have impacted the outcome significantly.

One other possible research area would be related to Bergin and Landy's "straw framed in tree bark" experiment. Taking the same scene from the four sensors and filtering it down to pure contrast information would allow a more quantitative and exact analysis of local texture segmentation on the target boundary. Another avenue to be explored would

involve manually manipulating the number of distractors in a natural NVD scene under a wide range of illumination and thermal conditions. While this method would be labor intensive (e.g., physically moving more trucks onto a field with a tank embedded as a target, under various illumination and temperature conditions) it would possibly allow determination of 'parallel' visual processes (in the style of Wolfe's 'Canal World') while also providing a detailed look at the wide spectrum of performance that can be expected from fusion and coloring devices.

Regardless of which search paradigm is chosen for future research on imagery from the four displays, a complete data set representative of the spectrum of fusion and coloring algorithms as well as the full range of IR and I^2 capabilities (which was not available for this thesis) is needed to completely assess human visual performance tasks with this technology.

APPENDIX A. IMAGE FILE CODING PROCEDURES

DOS filename limit: _____ _____ _____ _____ _____ _____ _____

1 2 3 4 5 6 7

Character codes:

1) Sensor

a) **i** = IR

b) **n** = I^2

c) **f** = fused

2) Fused?

a) **0** = not fused

b) **1** = fused color

c) **2** = fused monochrome

3 -5) Three letter description or acronym

e.g., trk for truck

6) Location of target

a) **0** = no target

b) **1** = original pos

c) **2** = a coherent position

d) **3** = another coherent position

7) Algorithm / producer

a) **a** = army fusion

b) **n** = nrl

c) **o** = original single-band image

APPENDIX B. REPRINT PERMISSION

Naval Research Laboratory
Code 5636
Washington D.C., 20378

September 12, 1996

Captain Matthew T. Sampson, USMC
Naval Postgraduate School, Monterey CA

Dear Captain Sampson,

You have my permission to use the tutorial viewgraphs that we previously supplied and the associated processed images for official use as part of your thesis.

Sincerely,

Dean Scribner

Dean Scribner, Ph.D.
Research Physicist

UNITED STATES MARINE CORPS
MARINE AVIATION WEAPONS AND TACTICS SQUADRON-1
Box 99200
Yuma, Arizona 85369-9200

IN REPLY REFER TO:
1500
NVD amso
18 Sep 96

From: Commanding Officer, Marine Aviation Weapons and Tactics Squadron One, Box 99200, Yuma, AZ 85369-9200

To Naval Postgraduate School, Operations Research Department, Glasgow Hall, Monterey, CA 93943-5219

Subj: **AUTHORIZATION FOR REPRINT OF MAWTS-1 NVD MANUAL DIAGRAMS**

Ref: (a) NPS Operations Research Department (CAPT Sampson) ltr request of 11 Sep 96

1. Per the reference request, you are authorized to use reprint diagrams from the MAWTS-1 Assault Support and TACAIR NVD manuals to support your Naval Postgraduate School Masters thesis. The subject diagrams used in the NVD manuals were originally adapted from various DOD technical reports with distribution authorized to U.S. Government Agencies and their contractors for administrative and operational use. Any additional uses or intended publication of these diagrams will require further DOD authorization. Please ensure that MAWTS-1 is included on the distribution list for this thesis project.

A. E. Kolkmeyer
By direction

Copy to:
LT Schutt (MAWTS-1 AMSO)

APPENDIX C. POWER AND SELECTION OF SAMPLE SIZE

Experimental Design Procedures for the Behavioral Sciences (Kirk), section 1.3 pp 9-11:

Dependent and Independent Variables have been determined.

Dependent: Reaction Time

Independent:

Sensor - IR, I^2, fused, fused color

Scene - Tower, Truck, Rectangle

Position - Target, Distractor

* Def: **Type I error** - *type I error* (α) is committed when the null hypothesis is rejected when it is in fact true.

* Def: **Type II error** - *type II error* (β) is committed when the null hypothesis is accepted when the alternative hypothesis is true (the null is false).

* Def: **Power** - the *power* of a research methodology is the probability of rejecting the null hypothesis when the alternative hypothesis is true or 1-[probability of committing a type II error (β)].

Sample size needs to be determined and five factors need to be considered in that determination:

1) The minimum treatment effects to be detected ($\mu_j - \mu$)

2) The number of treatment levels (k).

3) Population error variance (σ^2_ϵ).

4) Probability of making a type I error or significance level(α).

5) Probability of making a type II error (β) or power ($1-\beta$).

* population error variance (σ^2_ϵ) and the grand mean of the treatment effects (μ) are usually unknown but estimates using pilot studies can be made (Pilot study completed Nov 1995).

Below is the formula for the non-centrality coefficient (ϕ) used in Tang's method

for determining the Power and ultimately the correct sample (# subjects) size for the desired power. Tables of the power function for analysis of variance were available in Kirk (1968). These tables, based on the non-central F distribution with degrees freedom of the numerator (grand mean of the treatments estimated) $\upsilon_1 = k-1$ and degrees freedom of the denominator (individual treatment means estimated) $\upsilon_2 = N-k$. Since the Population error variance (σ^2_ϵ) was known from the pilot study, this formula, the desired power (≥ 0.80) and the degrees freedom of the denominator were used to derive a ϕ from the table. This non-centrality coefficient was then input in to the following equation and and the sample size (n) was solved for.

$$\phi = \frac{\sqrt{\sum_{1}^{k} \frac{(\mu_j - \mu)^2}{k}}}{\frac{\sigma_\epsilon}{\sqrt{n}}}$$

Using the four sensors as treatments:

1) The minimum acceptable treatment effects squared [$(\mu_j - \mu)^2$]:

I^2	$(757 - 803)^2 = 2,116.0$
IR	$(846 - 803)^2 = 1,849.0$
Fused Monochrome	$(787 - 803)^2 = 256.0$
Fused Color	$(822 - 803)^2 = 361.0$

Total = 4,582.0

2) The number of treatment levels (k): 4

3) Population error variance (σ^2_ϵ): 93813.0 (Pilot: 70886.0)

$\sigma_\epsilon = 306.3$ (Pilot: 266.2)

4) Probability of making a type I error or significance level(α): 0.05

5) Initial (pilot) size of sample per treatment (n): 810 or 162 independent observations per subject.

6) Independent observations (N = k x n): 3,240

7) Degrees freedom of denominator (df = N - k): 3,236 or essentially ∞

8) Degrees freedom of numerator (df = k-1): 3

9) Desired power: ≥0.80

10) Non-centrality coefficient (ϕ) for *at least* 0.8 power derived from tables: 1.65

11) Calculated non-centrality coefficient (ϕ): 3.619 (Pilot: 1.502) therefore power by Tang's method is *at least* .8

In S-plus the formula for calculating power from the non-central F distribution is:

1-pf(qf(p, df1, df2), df1, df2, ncp=0)

where pf = probability density, qf = quantile desired, df1 = df numerator, df2 = df denominator and ncp = non-centrality parameter δ. The non-centrality parameter ϕ is transformed to δ by the following method described in Johnson & Kotz (1970, V2):

$$\delta = \phi^2(df1 + 1)$$

11) Resultant δ: 40.17

12) S-plus code "power<-(1-pf(qf(.95, 3, 3236), 3, 3236, ncp=40.17))" yeilded a power of:

power = 0.9999192

Using 16 combinations of sensor by position was not required since the data analysis using SAS indicated it was statistically insignificant (F=0.1303, Pr(F)=0.942), therefore the data was collapsed to 8 combinations of sensor by target/distractor and analyzed for sensor by scene by target/distractor, which is significant (F=2.15, Pr(F)=0.0447). The following were the inputs used in the methods:

1) The minimum acceptable treatment effects squared [$(\mu_j - \mu)^2$]:

I^2 RectangleTarget	$(761 - 803)^2 = 1{,}764.0$
IR RectangleTarget	$(736 - 803)^2 = 4{,}489.0$
Fused Mono. RectangleTarget	$(836 - 803)^2 = 1{,}089.0$
Fused Color RectangleTarget	$(757 - 803)^2 = 2{,}116.0$
I^2 Rectangle Distractor	$(829 - 803)^2 = 676.0$
IR Rectangle Distractor	$(846 - 803)^2 = 1{,}849.0$
Fused Mono.RectangleDistractor	$(897 - 803)^2 = 8{,}836.0$
Fused Color RectangleDistractor	$(874 - 803)^2 = 5{,}041.0$
I^2 Truck Target	$(690 - 803)^2 = 12{,}769$
IR Truck Target	$(583 - 803)^2 = 48{,}400.0$
Fused Mono. Truck Target	$(610 - 803)^2 = 37{,}249.0$
Fused Color Truck Target	$(611 - 803)^2 = 36{,}864.0$
I^2 Truck Distractor	$(712 - 803)^2 = 8{,}281.0$
IR Truck Distractor	$(753 - 803)^2 = 2{,}500.0$
Fused Mono.Truck Distractor	$(730 - 803)^2 = 5{,}329.0$
Fused Color Truck Distractor	$(746 - 803)^2 = 3{,}249.0$
I^2 Tower Target	$(608 - 803)^2 = 38{,}025.0$
IR Tower Target	$(897 - 803)^2 = 8{,}836.0$
Fused Mono. Tower Target	$(647 - 803)^2 = 24{,}336.0$
Fused Color Tower Target	$(733 - 803)^2 = 4{,}900.0$
I^2 Tower Distractor	$(939 - 803)^2 = 18{,}496.0$
IR Tower Distractor	$(1258 - 803)^2 = 207{,}025.0$
Fused Mono.Tower Distractor	$(999 - 803)^2 = 38{,}416.0$
Fused Color Tower Distractor	$(1209 - 803)^2 = 164{,}836.0$

Total = 685,371.0

2) The number of treatment levels (k): 24

3) Population error variance (σ^2_ϵ): 93813.0 (Pilot: 70886.0)

$\sigma_\epsilon = 306.3$ (Pilot: 266.2)

4) Probability of making a type I error or significance level(α): 0.05

5) Initial (pilot) size of sample per treatment (n): 135 or 27 independent observations per subject.

6) Independent observations (N = k x n): 3,240

7) Degrees freedom of denominator (df = N - k): 3,216 or essentially ∞

8) Degrees freedom of numerator (df = k-1): 23

9) Desired power: ≥0.80

10) Calculated non-centrality coefficient (ϕ): 6.410

In S-plus the formula for calculating power from the non-central F distribution is:

1-pf(qf(p, df1, df2), df1, df2, ncp=0)

where pf = probability density, qf = quantile desired, df1 = df numerator, df2 = df denominator and ncp = non-centrality parameter δ. The non-centrality parameter ϕ is transformed to δ by the following method described in Johnson & Kotz (1970, V2):

$$\delta = \phi^2(df1 + 1)$$

11) Resultant δ: 986.114

12) S-plus code "power<-(1-pf(qf(.95, 23, 3216), 23, 3216,ncp=986.114))" yeilded a power of:

power = 1

These findings are consistent with Tang's tables which show an increase in power as you increase the degrees freedom in the numerator (treatments) while keeping the degrees freedom in the denominator (independent trials) essentially the same. Adjustments to the parameters above to account for the repeated measures analysis did not impact the results shown.

APPENDIX D. SUBJECT DEMOGRAPHICS

Subject #	Age	Rank	Service	Aero-Adapt?	NVD User?	Sex	MOS
1	37	O-4	USMC	yes	yes	M	Pilot
2	28	O-3	USMC	no	no	M	Maintenance
3	28	O-3	USMC	yes	yes	M	Pilot
4	27	O-3	USN	yes	no	M	NFO
5	34	O-3	USN	no	no	M	Submarines
6	32	O-3	Argentine	no	no	M	Surface

APPENDIX E. DATA ANALYSIS TECHNIQUES

Text output files from the experiment were collected by the VRG software and collated for each subject. After enhancements such as block and session number were added to each trial, the data was saved as a spreadsheet. The following SAS code was used to perform the MANOVA and ANOVAs:

```
/*DATA test;*/
options linesize=75;
options pagesize=200;
title " Sensor data analysis MANOVA";
data one (keep = sensor  scene  position  producer subject  aeroadpt  vision
        nvduse  time reactime);
    infile "sensor.txt";
    input  sensor $  scene $  position $  producer $ trial $  subject $  aeroadpt $  vision
$  nvduse $  session $ block $  stimulus $ response $  reactime  ;
if (stimulus NE response) then accuracy = 0; else accuracy = 1;
/*
if (position NE "0") then target="Y";  else target="N";
*/
proc sort; by  sensor  scene  position  producer subject  aeroadpt  vision  nvduse  time;
proc transpose out=new;
 by  sensor  scene  position  producer subject  aeroadpt  vision nvduse  ;
 id time;
proc print;
proc anova;
 class sensor scene position  subject  aeroadpt nvduse session  stimulus ;
 model accuracy reactime = sensor scene position aeroadpt nvduse session
                    subject
                    sensor*scene scene*position
                    sensor*position scene*sensor*position
                    sensor*position scene*sensor*position
                    sensor*aeroadpt position*aeroadpt
                    sensor*nvduse position*nvduse /nouni;
```

```
manova h = sensor scene position aeroadpt nvduse session subject
        sensor*scene scene*position sensor*position
        scene*sensor*position sensor*position scene*sensor*position
        sensor*aeroadpt position*aeroadpt sensor*nvduse
        target*nvduse /printe printh;

run;
```

LIST OF REFERENCES

Adelson, Edward H. & Bergen, James R. (1991). Plenoptic function and elements of early vision. In Landy, Micheal S. & Movshon, Anthony J. (Eds.), Computational models of visual processing. Cambridge, MA: The MIT press.

Amick, Daniel J. & Walberg, Herbert J. (Eds.). (1975). Introductory multivariate analysis for educational, psychological and social research. Berkeley, CA: McCutchan.

Bergen, James R. & Landy, Michael S. (1991). Computational model of visual texture segmentation. In Landy, Micheal S. & Movshon, Anthony J. (Eds.), Computational models of visual processing. Cambridge, MA: The MIT press.

Biederman, Irving, Glass, Arnold L., Stacy Jr. & E. Webb (1973). Searching for objects in real world scenes. Journal of Experimental Psychology, 97, 1.

Essock, E.A. (1992). An essay on texture: The extraction of stimulus structure from the visual image. In B. Burns (Ed.), Precepts, Concepts and Categories: The Representation and Processing of Visual Information. North-Holland Press, Elsevier Science Publishers, Amsterdam, Holland.

Hartline, P.H. and Newman, E.A. (1982). The infrared "vision" of snakes. Scientific American, 246, 3.

Hays, William L. (1988). Statistics (4th ed.). San Francisco, CA: Holt, Rinehart, Winston.

Julesz, B. (1984). A brief outline of the texton theory of human vision. Trends in Neuroscience, 7.

Kirk, R. E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole.

Krebs, W. (1994). A solid state visible/LWIR color night vision system to improve situational awareness and tactical efficiency. U.S. Navy, Office of Naval Research White Paper.

MAWTS-1 (1994, May). Tactical jet night vision device manual. (Available from [Commanding Officer, MAWTS-1, Box 99200, Yuma, AZ 85369-9200])

MAWTS-1 (1995, October). Helicopter night vision device manual. (Available from [Commanding Officer, MAWTS-1, Box 99200, Yuma, AZ 85369-9200])

MORS (1994). Military operations research analyst's handbook. (Available from [Military Operations Research Society, 101 South Whiting St., Alexandria, VA])

OPTEVFOR (1993, May). Tactics guide tz-1031-03-93 for A-6E night attack. (Available from [Commander, Operational Test and Evaluation Force, 7970 Diven St., Norfolk, VA 23505-1498])

Scribner, D., Satyshur, M., Shuler, J. & Kruer, M. (1996). Infrared Color Vision. Manuscript Submitted for Publication.

Sekuler, R. and Blake, R. (1990). Perception (2nd ed.). New York: McGraw-Hill.

SPIE (1987). Infrared sensors and sensor fusion. SPIE, 782.

Texas Instruments and U.S. Army (1993). Assessment of image fusion in a night pilotage system, NATO AC/243 Panel 3/4 Symposium on Multisensors and Sensor Fusion.

Treisman, A. (1985). Preattentive processing in vision. Computer Vision Graphics, & Image Processing, 31.

Treisman, A. (1986). Properties, parts and objects. In K. Boff, L. Kaufman & J. Thomas (Eds.), Handbook of perception and human performance. New York: Wiley.

Wolfe, J. M. (1994a). Visual search in continuous, naturalistic stimuli. Vision Research, 34.

Wolfe, J. M. (1994b). Guided Search 2.0, A Revised Model of Visual Search. Psychonomic Bulletin and Reveiw, 1 (2).

INITIAL DISTRIBUTION LIST

No. Copies

1. Defense Technical Information Center 2
   8725 John J. Kingman Road, Ste 0944
   Ft. Belvoir, VA 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Road
   Monterey, CA 93943-5101

3. Director,Training and Education 1
   MCCDC, Code C46
   1019 Elliot Rd.
   Quantico, VA 22134-5027

4. Director, Marine Corps Research Center 2
   MCCDC, Code C40RC
   2040 Broadway St.
   Quantico, VA 22134-5107

5. Director, Studies and Analysis Division 1
   MCCDC, Code C45
   3300 Russell Road
   Quantico, VA 22134-5130

6. W. K. Krebs, Code OR/Kw 10
   Naval Postgraduate School
   Monterey, CA 93943-5000

7. M. J. Youngren, Code OR/Ym 2
   Naval Postgraduate School
   Monterey, CA 93943-5000

8. T. M. Halwachs, Code OR/Ht 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

9. R. R. Read, Code OR/Rr 1
Naval Postgraduate School
Monterey, CA 93943-5000

10. W. Kemple, Code OR/Ke 1
Naval Postgraduate School
Monterey, CA 93943-5000

11. Naval Research Laboratory 1
Attn: D. Scribner, Code 5636
Washington, DC 20378

12. NAMRL 1
ATTN: CDR Still
51 Hovey Rd.
Pensacola, FL 32508-1046

13. Commanding Officer 2
MAWTS-1
Box 99200
Yuma, AZ 85369-9200

14. Capt Matthew T. Sampson USMC 2
MAWTS-1, ADT&E
Box 99200
Yuma, AZ 85369-9200